

**TRANSMITTAL LETTER TO THE
UNITED STATES
DESIGNATED/ELECTED OFFICE
(DO/EO/US) CONCERNING A FILING
UNDER 35 U.S.C. 371**

INTERNATIONAL APPLICATION NO.
PCT/FR98/01460INTERNATIONAL FILING DATE
July 7, 1998PRIORITY DATE CLAIMED
July 7, 1997

TITLE OF INVENTION
RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS
AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES

APPLICANT(S) FOR DO/EO/US

Glauca PARAHNOS-BACCALA, Florence KOMURIAN-PRADEL, Frederic BEDIN, Mireille SODOYER, Catherine OTT, Francois MALLET, Herve PERRON and Bernard MANDRAND

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. ☐ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
 - a. ☐ is transmitted herewith (required only if not transmitted by the International Bureau).
 - b. ☒ has been transmitted by the International Bureau.
 - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US)
6. ☒ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
 - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
 - b. ☐ have been transmitted by the International Bureau.
 - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
 - d. ☐ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371 (c)(5)).

Items 11. to 16. below concern other document(s) or information included:

11. ☒ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☒ A FIRST preliminary amendment.
☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A small entity statement.
16. ☐ Other items or information:

U.S. APPLICATION NO. (if known, see 37 C.F.R. 1.5) **09/319136**

INTERNATIONAL APPLICATION NO.
PCT/FR98/01460

ATTORNEY'S DOCKET NUMBER
103514

17. ☒ The following fees are submitted:

Basic National fee (37 CFR 1.492(a)(1)-(5)):

Search Report has been prepared by the EPO or JPO.....\$840.00

International preliminary examination fee paid to USPTO
(37 CFR 1.482).....\$670.00

No international preliminary examination fee paid to USPTO
(37 CFR 1.482) but international search fee paid to USPTO
(37 CFR 1.445(a)(2)).....\$760.00

Neither international preliminary examination fee (37 CFR
1.482) nor international search fee (37 CFR 1.445(a)(2))
paid to USPTO.....\$970.00

International preliminary examination fee paid to USPTO
(37 CFR 1.482) and all claims satisfied provisions of PCT
Article 33(2)-(4).....\$ 96.00

ENTER APPROPRIATE BASIC FEE AMOUNT =

\$840.00

\$840.00

Surcharge of \$130.00 for furnishing the oath or declaration later than
☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR
1.492(e)).

\$

Claims	Number Filed	Number Extra	Rate
--------	--------------	--------------	------

Total Claims	26- 20 =	---	X \$ 18.00
--------------	----------	-----	------------

\$108.00

Independent Claims	14- 3 =	---	X \$ 78.00
--------------------	---------	-----	------------

\$858.00

Multiple dependent claim(s)(if applicable)			+ \$260.00
--------------------------------------------	--	--	------------

\$

TOTAL OF ABOVE CALCULATIONS =

\$1,806.00

Reduction by 1/2 for filing by small entity, if applicable. Verified Small
Entity Statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28). -

\$---

SUBTOTAL =

\$1,806.00

Processing fee of \$130.00 for furnishing the English translation later
than ☐ 20 ☐ 30 month from the earliest claimed priority date (37 CFR
1.492(f)). +

\$---

TOTAL NATIONAL FEE =

\$1,806.00

Amount to be
refunded \$

Charged \$

- a. ☒ Check No. 100859 in the amount of \$1,806.00 to cover the above fees is enclosed.
- b. ☐ Please charge my Deposit Account No. _____ in the amount of \$_____ to cover the above fees. A duplicate copy of this sheet is enclosed.
- c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment, to Deposit Account No. 15-0461. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

OLIFF & BERRIDGE, PLC
P.O. Box 19928
Alexandria, Virginia 22320

NAME: William P. Berridge
REGISTRATION NUMBER: 30,024

NAME: Melanie L. Mealy
REGISTRATION NUMBER: 40,085

09/319156

PATENT APPLICATION

Rec'd PCT/PTO 02 JUN 1999

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of

Glaucia PARAHNOS-BACCALA, Florence KOMURIAN-
PRADEL, Frederic BEDIN, Mireille SODOYER, Catherine
OTT, Francois MALLET, Herve PERRON and Bernard
MANDRAND

Application No.: U.S. National Stage of
PCT/FR98/01460

Filed: June 2, 1999

Docket No.: 103514

For: RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN
PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR
RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND
THERAPEUTIC USES

PRELIMINARY AMENDMENT

Assistant Commissioner of Patents
Washington, D. C. 20231

Sir:

Prior to initial examination, please amend the above-identified application as follows:

IN THE CLAIMS:

Please amend claims 8, 13, 18, 20, 22, 23, 25 and 26 as follows:

Claim 8, line 4, delete "at nucleotide [sic]".

Claim 13, lines 1-2, change "any one of the preceding claims" to --claim 3--.

Claim 18, line 5, change "any one of claims 14 to 17" to --claim 14--.

Claim 20, line 7, change "any one of claims 8 to 11" to --claim 8--.

Claim 22, lines 3-5, change "any one of claims 1 to 7 or a fragment according to any
one of claims 14 to 17" to --claim 1--.

Claim 23, lines 2-3, change "any one of claims 14 to 17" to --claim 14--.

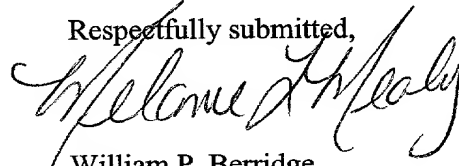
Claim 25, lines 5-6, change "any one of claims 14 to 17" to --claim 14--.

Claim 26, lines 7-8, change "any one of claims 14 to 17" to --claim 14--.

REMARKS

Claims 1-26 are pending. By this Preliminary Amendment, claim 8 is amended to eliminate unnecessary language and claims 13, 18, 20, 22, 23, 25 and 26 are amended to eliminate multiple dependency. Prompt and favorable examination is respectfully requested.

Respectfully submitted,



William P. Berridge
Registration No. 30,024

Melanie L. Mealy
Registration No. 40,085

WPB:MLM/sfh

OLIFF & BERRIDGE, PLC
P.O. Box 19928
Alexandria, Virginia 22320
Telephone: (703) 836-6400

RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS,
IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR
RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND
THERAPEUTIC USES

5

Multiple sclerosis (MS) is a demyelinating disease of the central nervous system (CNS) of which the complete cause still remains unknown.

10 Numerous studies have supported the hypothesis for a viral etiology of the disease, but none of the known viruses tested has proved to be the causative agent tested for: a review of the viruses tested for in MS for many years has been carried out by E. Norrby and R.T. Johnson.

15 Recently, a retrovirus, different from the known human retroviruses, was isolated from patients suffering from MS. The authors were able to show that this retrovirus could be transmitted in vitro, that patients suffering from MS produced antibodies capable of recognizing proteins associated with the infection
20 of the leptomeningeal cells by this retrovirus, and that the expression of the latter could be greatly stimulated by the immediate-early genes of some herpesviruses.

25 All these results argue in favor of the role in MS of at least one unknown retrovirus or of a virus having a reverse transcriptase (RT) activity which is detectable by the method published by H. Perron and termed "LM7-type RT" activity.

30 The studies by the applicant have made it possible to obtain two continuous cell lines infected with natural isolates obtained from two different patients suffering from MS, by a culture method as described in the document WO-A-93 20188, whose content
35 is incorporated by reference into the present description. These two lines derived from cells of human choroid plexus, called LM7PC and PLI-2, were deposited at the E.C.A.C.C. on 22 July 1992 and 8 January 1993, respectively, under numbers 92 072201

and 93 010817, in accordance with the provisions of the Treaty of Budapest. Moreover, the viral isolates possessing an LM7-type RT activity have also been deposited at the E.C.A.C.C. under the overall name of "strains". The "strain" or isolate harbored by the PLI-2 line, called POL-2, was deposited at the E.C.A.C.C. on 22 July 1992 under No. V92072202. The "strain" or isolate harbored by the LM7PC line, called MS7PG, was deposited at the E.C.A.C.C. on 8 January 1993 under No. V93010816.

Using the abovementioned cultures and isolates, characterized by biological and morphological criteria, efforts were then made to characterize the genetic material associated with the viral particles produced in these cultures.

The proportions of genome already characterized were used to develop molecular detection tests for the viral genome and immunoserological tests, using the amino acid sequences encoded by the nucleotide sequences of the viral genome, in order to detect the immune response directed against epitopes associated with the viral infection and/or expression.

These tools have already made it possible to confirm an association between MS and the expression of the sequences identified in the patents cited further on. However, the viral system discovered by the applicant is related to a complex retroviral system. Indeed, the sequences which are found to be encapsidated in the extracellular viral particles produced by the different cultures of cells of patients suffering from MS show clearly that there is co-encapsulation of retroviral genomes which are related but different from the "wild-type" retroviral genome which produces the infectious viral particles. This phenomenon was observed between replicative retroviruses and endogenous retroviruses belonging to the same family, or even heterologous retroviruses. The concept of endogenous retrovirus is very important in the context of our discovery because, in the case of

MSRV-1, it has been observed that endogenous retroviral sequences comprising sequences homologous to the MSRV-1 genome exist in normal human DNA. The existence of endogenous retroviral elements (ERV) related to MSRV-1 through all or part of their genome explains the fact that the expression of the MSRV-1 retrovirus in human cells can interact with related endogenous sequences. These interactions are found in the case of pathogenic and/or infectious endogenous retroviruses (for example some ecotropic strains of the Murine Leukemia virus), in the case of exogenous retroviruses whose nucleotide sequence may be found partially or completely in the form of ERVs, in the genome of the host animal (e.g. mouse mammary tumor exogenous virus transmitted via milk). These interactions consist mainly of (i) a transactivation or co-activation of ERVs by the replicative retrovirus, (ii) an "illegitimate" encapsidation of related RNAs of ERVs, or of ERVs - or even of cellular RNAs - simply possessing compatible encapsidation sequences, into the retroviral particles produced by the expression of the replicative strain, which are sometimes transmissible and sometimes with an inherent pathogenicity, and (iii) relatively high recombinations between the co-encapsidated genomes, in particular in the reverse transcription phases, which lead to the formation of hybrid genomes, which are sometimes transmissible and sometimes with an inherent pathogenicity.

Thus, (i) various MSRV-1-related sequences have been found in purified viral particles; (ii) molecular analysis of the various regions of the MSRV-1 retroviral genome should be carried out by systematically analyzing the co-encapsidated, interfering and/or recombinant sequences which are generated by the infection and/or expression of MSRV-1; furthermore, some clones may have portions of defective sequences produced by the retroviral replication and the template and/or transcription errors caused by reverse transcriptase; (iii) the families of sequences

related to the same retroviral genomic region are the supports for an overall diagnostic detection which may be optimized by the identification of invariable regions among the clones expressed and by the

5 identification of reading frames responsible for the production of antigenic and/or pathogenic polypeptides which may only be produced by a portion, or even only one, of the clones expressed and under these conditions, the systematic analysis of the clones

10 expressed in one region of a given gene makes it possible to evaluate the frequency of variation and/or recombination of the MSRV-1 genome in this region and to define the optimum sequences for the applications, in particular the diagnostic applications; (iv) the

15 pathology caused by a retrovirus such as MRSV-1 may be a direct effect of its expression and of the proteins or peptides produced as a result, but also an effect of the activation, encapsidation, recombination of related or heterologous genomes and proteins or peptides

20 produced as a result; thus, these genomes associated with the expression and/or infection by MSRV-1 are an integral part of the potential pathogenicity of this virus and therefore constitute diagnostic detection supports and particular therapeutic targets. Likewise,

25 any agent which is associated with, or which is a cofactor for these interactions responsible for the pathogenicity in question, such as MSRV-2 or the gliotoxic factor described in the patent application published under the No. FR-2,716,198, can participate

30 in the development of an overall and very effective strategy for therapeutic diagnosis, prognosis, monitoring and/or integrated therapy for MS in particular, but also for any other disease associated with the same agents.

35 In this context, a parallel discovery has been made in another autoimmune disease, rheumatoid arthritis (RA), which has been described in the French patent application filed under the No. 95 02960. This discovery shows that, by applying methodological

approaches similar to those which were used in the studies by the applicant on MS, it has been possible to identify a retrovirus expressed in RA which shares the sequences described for MSRV-1 in MS and also the co-existence of an MSRV-2-associated sequence which is also described in MS. As regards MSRV-1, the sequences commonly detected in MS and RA relate to the *pol* and *gag* genes. On the basis of current knowledge, it is possible to combine the *gag* and *pol* sequences described with the MSRV-1 strains expressed in these two diseases.

The present patent application has as its object various results, supplementary in relation to those already protected by the French patent applications:

- No. 92/04322 of 03.04.1992, published under No. 2,689,519;
- No. 92/13447 of 03.11.1992, published under No. 2,689,521;
- 20 - No. 92/13443 of 03.11.1992, published under No. 2,689,520;
- No. 94/01529 of 04.02.1994, published under No. 2,715,936;
- No. 94/01531 of 04.02.1994, published under No. 2,715,939;
- 25 - No. 94/01530 of 04.02.1994, published under No. 2,715,936;
- No. 94/01532 of 04.02.1994, published under No. 2,715,937;
- 30 - No. 94/14322 of 24.11.1994, published under No. 2,727,428;
- No. 94/15810 of 23.12.1994, published under No. 2,728,585;

and

- 35 - Patent Application WO 97/06260.

The present invention relates, first of all, to a nucleic material, which may consist of a retroviral material, in isolated or purified state, which may be understood or characterized in various ways:

- it comprises a nucleotide sequence chosen from the group which consists of (i) the sequences
SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117,
SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130,
5 SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences
complementary to sequences (i); and (iii) the sequences
equivalent to sequences (i) or (ii), in particular the
sequences having, for every series of 100 contiguous
monomers, at least 50%, and preferentially at least 70%
10 homology with sequences (i) or (ii) respectively;

- it encodes a polypeptide having, for every
contiguous series of at least 30 amino acids, at least
50%, and preferably at least 70% homology with a
peptide sequence chosen from the group which consists
15 of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118,
SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;

- its pol gene comprises a nucleotide sequence
identical or equivalent to a sequence chosen from the
group which consists of SEQ ID NO: 112, SEQ ID NO: 124
20 and their complementary sequences;

- the 5' end of its pol gene starts at
nucleotide 1419 of SEQ ID NO: 130;

- its pol gene encodes a polypeptide having,
for every contiguous series of at least 30 amino acids,
25 at least 50%, and preferably at least 70% homology with
the peptide sequence SEQ ID NO: 113;

- the 3' end of its gag gene ends at nucleotide
1418 of SEQ ID NO: 130;

- its env gene comprises a nucleotide sequence
30 identical or equivalent to a sequence chosen from the
group which consists of SEQ ID NO: 117, and its
complementary sequences;

- its env gene comprises a nucleotide sequence
which starts at nucleotide 1 of SEQ ID NO: 117 and ends
35 at nucleotide at nucleotide [sic] 233 of SEQ ID NO:
114;

- its env gene encodes a polypeptide having,
for every contiguous series of at least 30 amino acids,

at least 50%, and preferably at least 70% homology with the sequence SEQ ID NO: 118;

- the U3R region of its 3' LTR comprises a nucleotide sequence which ends at nucleotide 617 of
5 SEQ ID NO: 114;

- the RU5 region of its 5' LTR comprises a nucleotide sequence which starts at nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID NO: 141 or SEQ ID NO: 142;

10 - a retroviral nucleic material comprising a sequence which starts at nucleotide 755 of SEQ ID NO: 120 and which ends at nucleotide 617 of SEQ ID NO: 114;

- the retroviral nucleic material as defined above is in particular associated with at least one
15 autoimmune disease such as multiple sclerosis or rheumatoid arthritis.

The invention also relates to a nucleotide fragment which corresponds to at least one of the following definitions:

20 - it comprises or consists of a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii)
25 the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i)
30 or (ii) respectively;

- it comprises or consists of a nucleotide sequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a
35 peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

Other subjects of the present invention are the following:

- a nucleic probe for the detection of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, capable of hybridizing specifically with any fragment defined above and
5 belonging to the genome of said retrovirus; it advantageously possesses from 10 to 100 nucleotides, preferably from 10 to 30 nucleotides;

- a primer for the amplification, by polymerization, of an RNA or of a DNA of a retrovirus
10 associated with multiple sclerosis and/or rheumatoid arthritis, which comprises a nucleotide sequence identical or equivalent to at least a portion of the nucleotide sequence of a fragment defined above, in particular a nucleotide sequence having, for every
15 series of 10 contiguous monomers, at least 50%, preferably at least 70% homology with at least said portion of said fragment; preferably the nucleotide sequence of a primer of the invention is chosen from
SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122,
20 SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127,
SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, and
SEQ ID NO: 133;

- an RNA or a DNA, and in particular a replication and/or expression vector, comprising a
25 genomic fragment of the nucleic material or a fragment defined above;

- a peptide encoded by any open reading frame belonging to a nucleotide fragment defined above, in particular a polypeptide, for example oligopeptide
30 forming or comprising an antigenic determinant recognized by sera of patients infected with the MSRV-1 virus, and/or in whom the MSRV-1 virus has been reactivated; a preferential peptide comprises a sequence identical, partially or completely, or
35 equivalent to a sequence chosen from SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;

- a diagnostic, prophylactic or therapeutic composition, in particular for inhibiting the

expression of at least one retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, comprising a nucleotide fragment defined above;

- a method for detecting a retrovirus
5 associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, comprising the steps consisting of bringing an RNA and/or a DNA assumed to belong to or obtained from said retrovirus, or their complementary RNA and/or DNA, into contact with a
10 composition comprising a nucleotide fragment defined above.

Before detailing the invention, various terms used in the description and the claims are now defined:

- strain or isolate is understood to mean any
15 infectious and/or pathogenic biological fraction containing, for example, viruses and/or bacteria and/or parasites, generating a pathogenic and/or antigenic power, harbored by a culture or a live host; by way of example, a viral strain according to the preceding
20 definition may contain a co-infectious agent, for example a pathogenic protist,

- the term "MSRV" used in the present description designates any pathogenic and/or infectious agent, as associated with MS, in particular a viral
25 species, the attenuated strains of said viral species, or the interfering defective particles or particles containing co-encapsidated genomes or alternatively genomes recombined with a portion of the MSRV-1 genome, which are derived from this species. It is known that
30 viruses and particularly viruses containing RNA exhibit variability, following in particular relatively high rates of spontaneous mutation, which will be taken into account below to define the concept of equivalence,

- human virus is understood to mean a virus
35 capable of infecting or of being harbored by human beings,

- given all the natural or induced variations and/or recombination which may be encountered in practice in the present invention, the objects thereof,

defined above and in the claims, have been expressed by comprising the equivalents or derivatives of the various biological materials defined below, in particular homologous nucleotide or peptide sequences,

5 - the variant of a virus or of a pathogenic and/or infectious agent according to the invention comprises at least one antigen recognized by at least one antibody directed against at least one corresponding antigen of said virus and/or of said
10 pathogenic and/or infectious agent, and/or a genome in which any portion is detected by at least one hybridization probe, and/or at least one nucleotide amplification primer specific for said virus and/or pathogenic and/or infectious agent, under defined
15 hybridization conditions well known to persons skilled in the art,

 - according to the invention, a nucleotide fragment or an oligonucleotide or a polynucleotide is a stretch of monomers, or a biopolymer, characterized by
20 the informational sequence of the natural nucleic acids, which is capable of hybridizing to any other nucleotide fragment under predefined conditions, it being possible for the stretch to contain monomers of different chemical structures and to be obtained from a
25 natural nucleic acid molecule and/or by genetic recombination and/or by chemical synthesis; a nucleotide fragment may be identical to a genomic fragment of the MSRV-1 virus considered by the present invention, in particular a gene of the latter, for
30 example pol or env in the case of said virus;

 - thus, a monomer may be a natural nucleic acid nucleotide in which the constituent components are a sugar, a phosphate group and a nitrogen base; in RNA, the sugar is ribose; in DNA, the sugar is 2-deoxy-
35 ribose; depending on whether DNA or RNA is involved, the nitrogen base is chosen from adenine, guanine, uracil, cytosine, thymine; or the nucleotide may be modified in at least one of the three constituent components; by way of example, the modification may

occur at the level of the bases, generating modified bases such as inosine, 5-methyl-deoxycytidine, deoxyuridine, 5-dimethylamineodeoxyuridine [sic], 2,6-diamineopurine [sic], 5-bromodeoxyuridine and any
5 other modified base promoting hybridization; at the level of the sugar, the modification may consist in the replacement of at least one deoxyribose with a polyamide, and at the level of the phosphate group, the modification may consist in its replacement with
10 esters, in particular chosen from the esters of diphosphate, of alkyl and arylphosphonate and of phosphorothioate,

- "informational sequence" is understood to mean any ordered series of monomers, whose chemical
15 nature and in which the order in a reference direction, constitute or otherwise a functional information of the same quality as that for the natural nucleic acids,

- hybridization is understood to mean the process during which, under appropriate operating
20 conditions, two nucleotide fragments, having sufficiently complementary sequences, become annealed to form a complex, in particular a double or triple, structure, preferably in helical form,

- a probe comprises a nucleotide fragment
25 synthesized by the chemical route or obtained by digestion or enzymatic cleavage of a longer nucleotide fragment, comprising at least six monomers, advantageously from 10 to 100 monomers, preferably 10 to 30 monomers, and possessing a hybridization
30 specificity under defined conditions; preferably, a probe possessing less than 10 monomers is not used alone, but is used in the presence of other probes which are equally short in length or otherwise; under certain specific conditions, it may be useful to use
35 probes which are greater than 100 monomers in size; a probe may be used in particular for diagnostic purposes, and it may be, for example, capture and/or detection probes,

- the capture probe may be immobilized on a solid support by any appropriate means, that is to say directly or indirectly, for example by covalent bonding or passive adsorption,

5 - the detection probe may be labeled by means of a marker chosen in particular from radioactive isotopes, enzymes chosen in particular from peroxidase and alkaline phosphatase and those capable of hydrolyzing a chromogenic, fluorigenic or luminescent
10 substrate, chromophoric chemical compounds, chromogenic, fluorigenic or luminescent compounds, analogs of nucleotide bases, and biotin,

15 - the probes used for diagnostic purposes of the invention may be used in all known hybridization techniques, and in particular the so-called "DOT-BLOT" technique, "SOUTHERN BLOT" technique, "NORTHERN BLOT" technique which is a technique identical to the "SOUTHERN BLOT" technique but which uses RNA as target, the SANDWICH technique; advantageously, the SANDWICH
20 technique is used in the present invention, comprising a specific capture probe and/or a specific detection probe, it being understood that the capture probe and the detection probe must have a nucleotide sequence which is at least partially different,

25 - any probe according to the present invention may hybridize in vivo or in vitro with the RNA and/or with the DNA, in order to block the replication, in particular translation and/or transcription, phenomena and/or to degrade said DNA and/or RNA,

30 - a primer is a probe comprising at least six monomers, and advantageously from 10 to 30 monomers, possessing hybridization specificity under defined conditions, for the initiation of an enzymatic polymerization, for example in an amplification
35 technique such as PCR (Polymerase Chain Reaction), in an extension method such as sequencing, in a reverse transcription method and the like,

- two nucleotide or peptide sequences are said to be equivalent or derived with respect to each other,

or with respect to a reference sequence, if functionally the corresponding biopolymers can play substantially the same role, without being identical, in relation to the application or use considered, or in
5 the technique in which they are involved; particularly equivalent are two sequences obtained because of the natural variability, in particular spontaneous mutation, of the species from which they were identified, or induced mutation, as well as two
10 homologous sequences, the homology being defined below,

- "variability" is understood to mean any spontaneous or induced modification of a sequence, in particular by substitution, and/or insertion, and/or deletion of nucleotides and/or of nucleotide fragments,
15 and/or extension and/or shortening of the sequence at least at one of the ends; a nonnatural variability may result from the genetic engineering techniques used, for example from the choice of the degenerate or nondegenerate synthetic primers selected to amplify a
20 nucleic acid; this variability may result in modifications of any starting sequence, considered as a reference, and which may be expressed by a degree of homology with respect to said reference sequence,

- homology characterizes the degree of identity
25 of two compared nucleotide or peptide fragments; it is measured by the percentage identity which is in particular determined by direct comparison of nucleotide or peptide sequences, with respect to reference nucleotide or peptide sequences,

- any nucleotide fragment is said to be
30 equivalent to or derived from a reference fragment if it has a nucleotide sequence equivalent to the sequence of the reference fragment; according to the preceding definition, in particular equivalent to a reference
35 nucleotide fragment are:

(a) any fragment capable of hybridizing, at least partially, with the complementary to the reference fragment,

(b) any fragment whose alignment with the reference fragment leads to the identification of identical contiguous bases, in a greater number than with any other fragment obtained from another taxonomic group,

(c) any fragment resulting or capable of resulting from the natural variability of the species from which it is obtained,

(d) any fragment which may result from genetic engineering techniques applied to the reference fragment,

(e) any fragment, containing at least eight contiguous nucleotides, encoding a peptide homologous or identical to the peptide encoded by the reference fragment,

(f) any fragment different from the reference fragment through insertion, deletion, substitution of at least one monomer, extension, or shortening at least at one of its ends; for example, any fragment corresponding to the reference fragment, flanked at least at one of its ends by a nucleotide sequence not encoding a polypeptide,

- polypeptide is understood to mean in particular any peptide of at least two amino acids, in particular oligopeptide, protein, extracted, separated, or substantially isolated or synthesized, through the involvement of humans, in particular those obtained by chemical synthesis, or through expression in a recombinant organism,

- polypeptide partially encoded by a nucleotide fragment is understood to mean a polypeptide having at least three amino acids encoded by at least nine contiguous monomers included in said nucleotide fragment,

- an amino acid is said to be analogous to another amino acid when their respective physicochemical characteristics, such as polarity, hydrophobicity and/or basicity, and/or acidity, and/or

neutrality, are substantially the same; thus, a leucine is analogous to an isoleucine,

- any polypeptide is said to be equivalent to or derived from a reference polypeptide if the polypeptides compared have substantially the same properties, and in particular the same antigenic, immunological, enzymatic and/or molecular recognition properties; in particular equivalent to a reference polypeptide is:

10 (a) any polypeptide possessing a sequence in which at least one amino acid has been replaced by an analogous amino acid,

(b) any polypeptide having an equivalent peptide sequence, obtained by natural or induced variation of said reference polypeptide, and/or of the nucleotide fragment encoding said polypeptide,

(c) a mimotope of said reference polypeptide,

(d) any polypeptide from whose sequence one or more amino acids of the L series are replaced by an amino acid of the D series, and vice versa,

(e) any polypeptide into whose sequence a modification of the side chains of the amino acids has been introduced, such as for example an acetylation of the amine-containing functions, a carboxylation of the thiol functions, an esterification of the carboxyl functions,

(f) any polypeptide in whose sequence one or more peptide bonds have been modified, such as for example the carba, retro, inverso, retro-inverso, reduced, and methylene-oxy bonds,

(g) any polypeptide in which at least one antigen is recognized by an antibody directed against a reference polypeptide,

- the percentage identity characterizing the homology between two peptide fragments compared is according to the present invention at least 50% and preferably at least 70%.

Given that a virus possessing a reverse transcriptase enzymatic activity may be genetically

characterized both in RNA and DNA form, both the viral DNA and RNA will be mentioned in order to characterize the sequences relative to a virus possessing such a reverse transcriptase activity, termed MSRV-1 according to the present description.

The expressions of order which are used in the present description and the claims, such as "first nucleotide sequence", are not selected to express a particular order, but to define the invention more clearly.

Detection of a substance or agent is understood below to mean an identification, a quantification or a separation or isolation of said substance or of said agent.

The invention will be understood more clearly on reading the detailed description which follows which is made with reference to the appended figures in which:

Figure 1 represents the general structure of the proviral DNA and the genomic RNA of MSRV-1.

Figure 2 represents the nucleotide sequence of the clone called CL6-5' (SEQ ID NO: 112) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 3 represents the nucleotide sequence of the clone called CL6-3' (SEQ ID NO: 114) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 4 represents the nucleotide sequence of the clone called C15 (SEQ ID NO: 117) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 5 represents the nucleotide sequence of the clone called 5M6 (SEQ ID NO: 120) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 6 represents the nucleotide sequence of the clone called CL2 (SEQ ID NO: 130) and three

potential reading frames in amino acids presented under the nucleotide sequence.

Figure 7 represents three potential reading frames in amino acids expressed by pET28C-clone 2 and presented under the nucleotide sequence.

Figure 8 represents three potential reading frames in amino acids expressed by pET21C-clone 2 and presented under the nucleotide sequence.

Figure 9 represents the nucleotide sequence of the clone called LB13 (SEQ ID NO: 141) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 10 represents the nucleotide sequence of the clone called LA15 (SEQ ID NO: 142) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 11 represents the nucleotide sequence of the clone called LB16 (SEQ ID NO: 124) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 12 represents the promoter activity expressed in cpm/4 min of the U3R sequences subcloned from LTRs of different origins into the plasmid PCAT3. PCAT3 means plasmid alone, PCAT-PH74 means plasmid plus endogenous U3R clone expressed in the placenta, PCAT-cl6 means plasmid plus U3R clone amplified in the RNA of an MS plasma, PCAT-5M6 means plasmid plus U3R region amplified in the cellular DNA, "no plasmid" means absence of plasmid in the test.

Figure 13 represents the MSRV1 env and 3' LTR sequences. The horizontal arrows indicate the start of the env, U3 and R regions. In the env region, the signal peptide and the potential immunosuppressive region are underlined, the potential glycosilation sites are boxed and the potential cleavage sites are indicated by vertical arrows. In the U3R region: the regulatory element CAAT and the TATA Box are underlined, the "cap" site and the polyadenylation signal are also indicated.

Figure 14 represents the 5' LTR (RU5) region followed by a PBS site (primer binding site) complementary to the Trp tRNA and by a gag gene encoding a protein of about 487 amino acids. The amino acids conserved in the nucleocapsid are underlined twice. The amino acids defining the region of greatest homology in the capsid are in bold and underlined once. The / symbols in the amino acid sequence indicate variations observed depending on the clones and, in the nucleotide sequence, they indicate frame jumps in some clones. The boxed regions correspond to epitopes identified by peptide analysis of the C-terminal region.

Figure 15 represents the integrase region of MSRV1, the nucleotide sequence and the amino acid sequence deduced from the integrase region corresponding to clone 87-23. In Figure 15, // means a frame jump which has been suppressed in order to restore the potential ORF. The letters in underlined bold characters represent the conserved amino acids in the retroviral integrases.

Figure 16 describes the nucleotide and peptide sequences of clone B13 (identical to clone FBd13 described in previous applications) with indication of the ORFs and stop codons represented by a dot. The underlined region in bold represents the potential immunosuppressive domain. The single underlined domain represents the start of the 3' LTR.

EXAMPLE 1: PREPARATION OF A CL6-5' REGION ENCODING THE N-TERMINAL END OF INTEGRASE AND OF A CL6-3' REGION CONTAINING THE 3' TERMINAL SEQUENCE OF THE MSRV-1 GENOME

A 3' RACE was carried out on the total RNA extracted from plasma from a patient suffering from MS. A healthy control plasma, treated under the same conditions, was used as negative control. The synthesis of cDNA was carried out with an oligo dT primer identified by SEQ ID NO: 68 (5' GAC TCG CTG CAG ATC GAT

TTT TTT TTT TTT TTT T 3') and the reverse transcriptase "ExpandTM RT" from Boehringer according to the conditions recommended by the company. A PCR was carried out with the enzyme KlenTaq (Clontech) under the following conditions: 94°C 5 min then 93°C 1 min, 58°C 1 min, 68°C 3 min over 40 cycles and 68°C for 8 min, with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 69

10 5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';

- 3' primer, identified by SEQ ID NO: 68

A second so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used for the first PCR, using 10 µl of the amplification product derived from the first PCR.

Primers used for the seminested PCR:

- 5' primer, identified by SEQ ID NO: 70

20 5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';

- 3' primer, identified by SEQ ID NO: 68

The primers SEQ ID NO: 69 and SEQ ID NO: 70 are specific for the pol region of MRSV-1.

An amplification product of 1.9 Kb was obtained for the plasma of the MS patient. The corresponding fragment was not observed for the healthy control plasma. This amplification product was cloned in the following manner:

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit[®]. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCRTM VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 12°C. The next steps were carried out in accordance with the instructions for the TA Cloning kit[®] (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the

extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM[™] Ready Reaction AmpliTaq[®] FS, DyeDeoxy[™] Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The clone obtained contains a CL6-5' region encoding the N-terminal end of integrase and a CL6-3' region corresponding to the 3' terminal region of MSRV-1 and making it possible to define the end of the envelope (234 bp) and the U3 and R (401 bp) regions of the MSRV1 retrovirus.

The region corresponding to the N-terminal end of integrase is represented by its nucleotide sequence (SEQ ID NO: 112) in Figure 27. The three potential reading frames are presented by their amineo [sic] acid sequence under the nucleotide sequence, and the amineo [sic] acid sequence of the N-terminal end of integrase is identified by SEQ ID NO: 113.

The CL6-3' region is represented by its nucleotide sequence (SEQ ID NO: 114) in Figure 3. The three potential reading frames are presented by their amineo [sic] acid sequence under the nucleotide sequence. An amineo [sic] acid sequence corresponding to the C-terminal end of the MSRV-1 env protein is identified by SEQ ID NO: 115.

In order to evaluate the promoter activity of the LTR obtained from clone 6 (cl6), a test of promoter

activity using the enzyme CAT (chloramphenicol acetyl transferase) was carried out with the corresponding U3R region. In parallel, a clone containing the same U3R region of endogenous retroviral RNA expressed in normal placenta (PH74) and a clone (5M6) obtained from DNA were tested. The result presented in Figure 12 shows a very high promoter activity of the LTR derived from MS plasma (cl6) and a significantly much lower activity with the sequences of non-MS endogenous origin.

EXAMPLE 2: PREPARATION OF THE C15 CLONE CONTAINING THE REGION ENCODING A PORTION OF THE MSRV-1 RETROVIRUS ENVELOPE

A RT-PCR was carried out on the total RNA extracted from virions concentrated by ultracentrifugation of a synovocyte culture supernatant obtained from an MS patient. The synthesis of cDNA was carried out with an oligo dT primer and the reverse transcriptase "ExpandTM RT" from Boehringer according to the conditions recommended by the company. A PCR was carried out with the ExpandTM Long Template PCR System (Boehringer) under the following conditions: 94°C 5 min then 93°C 1 min, 60°C 1 min, 68°C 3 min over 40 cycles and 68°C for 8 min and with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 69
5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';
- 3' primer, identified by SEQ ID NO: 116
5' TGG GGT TCC ATT TGT AAG ACC ATC TGT AGC TT 3'

A second so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used for the first PCR (except that 30 cycles were used instead of 40), using 10 µl of the amplification product derived from the first PCR.

Primers used for the seminested PCR:

- 5' primer, identified by SEQ ID NO: 70

5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';

- 3' primer, identified by SEQ ID NO: 116

The primers SEQ ID NO: 69 and SEQ ID NO: 70 are specific for the pol region of MRSV-1. The primer SEQ ID NO: 116 is specific for the sequence FBd13 (also called B13) and is located in the conserved env region among the oncoretroviruses.

An amplification product of 1932 bp was obtained and cloned in the following manner:

10 the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit[®]. The various steps were carried out in accordance with the instructions for the TA Cloning kit[®] (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria
15 (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose
20 gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the SP6 promoter present on the cloning plasmid of the TA
25 cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM[™] Ready Reaction AmpliTaq[®] FS, DyeDeoxy[™] Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing
30 was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The C15 clone obtained contains a region corresponding to the region of the MRSV-1 envelope of
35 1481 bp.

The env region of the C15 clone is represented by its nucleotide sequence (SEQ ID NO: 117) in Figure 5. The three potential reading frames of this clone are presented by their amino acid sequence

under the nucleotide sequence. The reading frame corresponding to an MSRV-1 structural env protein is identified by SEQ ID NO: 118.

From the defined sequences obtained from clones
5 c16 and C15, it was possible to produce a plasmid construct encoding a complete envelope followed by the 3' LTR, as presented in Figure 13 with the corresponding reading frame.

10 **EXAMPLE 3:** PREPARATION OF A 5M6 CLONE CONTAINING THE SEQUENCES OF THE 3' TERMINAL REGION OF THE ENVELOPE, FOLLOWED BY THE MSRV-1 PROVIRAL TYPE U3, R AND U5 SEQUENCES

A monodirectional PCR was carried out on the
15 DNA extracted from immortalized B lymphocytes in culture from an MS patient. The PCR was carried out with ExpandTM Long Template PCR System (Boehringer) under the following conditions: 94°C 3 min then 93°C 1 min, 60°C 1 min, 68°C 3 min over 10 cycles, then 93°C
20 1 min, 60°C 1 min with 15 sec of extension at each cycle, 68°C 3 min over 35 cycles and 68°C for 7 min and with a final reaction volume of 50 µl.

The primer used for the PCR identified by SEQ
ID NO: 119 is 5' TCA AAA TCG AAG AGC TTT AGA CTT GCT
25 AAC CG 3';

The primers [sic] SEQ ID NO: 119 is specific for the env region of the C15 clone.

An amplification product of 1673 bp was obtained and cloned in the following manner:
30 the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit[®]. The various steps were carried out in accordance with the instructions for the TA Cloning kit[®] (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria
35 (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose

gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The 5M6 clone obtained contains a region corresponding to the 3' region of the MSRV-1 envelope of 492 bp followed by the regions U3, R and U5 (837 bp) of MSRV1.

The 5M6 clone is represented by its nucleotide sequence (SEQ ID NO: 120) in Figure 5. The three potential reading frames of this clone are presented by their amino acid sequence under the nucleotide sequence. The reading frame corresponding to the C-terminal end of the MSRV-1 env protein is identified by SEQ ID NO: 121.

EXAMPLE 4: PREPARATION OF THE LB16 CLONE CONTAINING THE REGION ENCODING THE MSRV-1 RETROVIRUS INTEGRASE

An RT-PCR was carried out on the total RNA treated with DNaseI and extracted from a choroid plexus obtained from an MS patient. The synthesis of cDNA was carried out with an oligo dT primer and the reverse transcriptase "Expand™ RT" from Boehringer according to the conditions recommended by the company. A "no RT" control was carried out in parallel on the same material. A PCR was carried out with Taq polymerase (Perkin Elmer) under the following conditions: 95°C 5 min, then 95°C 1 min, 55°C 1 min, 72°C 2 min over 35 cycles and 72°C for 8 min and with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 122

5' GGC ATT GAT AGC ACC CAT CAG 3';

- 3' primer, identified by SEQ ID NO: 123

5 5' CAT GTC ACC AGG GTG GAA TAG 3'

The primer SEQ ID NO: 122 is specific for the pol region of MSRV-1 and more precisely similar to the integrase region described above. The primer SEQ ID NO 123 was defined on sequences of the clones obtained during preliminary tests.

An amplification product of about 760 bp was obtained only in the test with RT and was cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit[®]. The various steps were carried out in accordance with the instructions for the TA Cloning kit[®] (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM[™] Ready Reaction AmpliTaq[®] FS, DyeDeoxy[™] Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The LB16 clone obtained contains the sequences corresponding to integrase. The nucleotide sequence of

this clone was identified by SEQ ID NO: 124 in Figure 11, three reading frames are determined.

EXAMPLE 5: PREPARATION OF A CLONE 2, CL2, CONTAINING IN 3' A PORTION HOMOLOGOUS TO THE POL GENE, CORRESPONDING TO THE PROTEASE GENE, AND TO THE GAG GENE (GM3) CORRESPONDING TO THE NUCLEOCAPSID, AND A NEW 5' CODING REGION, CORRESPONDING TO THE GAG GENE MORE SPECIFICALLY THE TEMPLATE AND THE CAPSID of MSRV-1.

10 A PCR amplification was carried out on the total RNA extracted from 100 µl of plasma from a patient suffering from MS. A water control, treated under the same conditions, was used as negative control. The synthesis of cDNA was carried out with 15 300 pmol of a random primer (GIBCO-BRL, France) and the reverse transcriptase "Expand RT" (BOEHRINGER MANNHEIM, France) according to the conditions recommended by the company. An amplification by PCR ("polymerase chain reaction") was carried out with the enzyme Taq 20 polymerase (Perkin Elmer, France) using 10 µl of cDNA under the following conditions: 94°C 2 min, 55°C 1 min, 72°C 2 min then 94°C 1 min, 55°C 1 min, 72°C 2 min over 30 cycles and 72°C for 7 min with a final reaction volume of 50 µl.

25 Primers used for the PCR amplification:
- 5' primer, identified by SEQ ID NO: 126
5' CGG ACA TCC AAA GTG ATG GGA AAC G 3';
- 3' primer, identified by SEQ ID NO: 127
5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

30 A second amplification by so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used during the first PCR, using 35 10 µl of the amplification product derived from the first PCR.

Primers used for the amplification by seminested PCR:
- 5' primer, identified by SEQ ID NO: 128
5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3';

- 3' primer, identified by SEQ ID NO: 129
5' TGG CTC TCA ATG GTC AAA CAT ACC CG 3'

The primers SEQ ID NO: [lacuna] and SEQ ID NO: [lacuna] are specific for the pol region, clone G+E+A, more specifically the E region: nucleotide position No. 423 to No. 448. The primers used in the 5' region were defined on sequences of clones obtained during preliminary tests.

An amplification product of 1511 bp was obtained from the RNA extracted from the plasma of an MS patient. The corresponding fragment was not observed for the water control. This amplification product was cloned in the following manner.

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kitTM. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCRTM VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning kit[®] (Invitrogen). The mixture was plated after transformation of the ligation into *E. coli* INVαF' bacteria. At the end of the procedure, the white colonies of recombinant bacteria were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "DNA minipreparation" procedure (17). The plasmid preparation of each recombinant colony was cut with the restriction enzyme EcoRI and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISMTM Ready Reaction Amplitaq[®] FS, DyeDeoxyTM Terminator" (Applied

Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

5 The clone obtained, called CL2, contains a C-terminal region similar to the 5' terminal region of the clones G+E+A of MSRV-1, which makes it possible to define the C-terminal region of the gag gene and a new region corresponding to the N-terminal region of the
10 MSRV-1 gag gene.

CL2 makes it possible to define a region of 1511 bp having an open reading frame in the N-terminal region of 1077 bp encoding 359 amino acids and a non-open reading frame of 454 bp corresponding to the
15 C-terminal region of the MSRV-1 gag gene.

The nucleotide sequence of CL2 is identified by SEQ ID NO: 130. It is represented in Figure 6 with the potential reading frames in amineo [sic] acid.

The 1077 bp fragment of CL2 encoding 359 amino
20 acids was amplified by PCR with the *Pwo* enzyme (5U/ μ l) (Boehringer Mannheim, France) using 1 μ l of the DNA minipreparation of clone 2 under the following conditions: 95°C 1 min, 60°C 1 min, 72°C 2 min over 25 cycles and with a final reaction volume of 50 μ l
25 with the aid of the primers:

- 5' primer (*Bam*HI), identified by SEQ ID NO: 132

5' TGC TGG AAT TCG GGA TCC TAG AAC GTA TTC 3' (30 mer),
and

- 3' primer (*Hind*III), identified by SEQ ID NO: 133

30 5 AGT TCT GCT CCG AAG CTT AGG CAG ACT TTT 3' (30 mer)
corresponding, respectively, to the nucleotide sequence of clone 2 at position -9 to 21 and 1066 to 1095.

The fragment obtained by PCR was linearized with *Bam*HI and *Hind*III and subcloned into the
35 expression vectors pET28C and pET21C (NOVAGEN) linearized with *Bam*HI and *Hind*III. The sequencing of the DNA of the 1077 bp fragment of clone 2 in the two expression vectors was carried out according to the method recommended for the use of the sequencing kit

"PRISMTM Ready Reaction Amplitaq[®] FS, DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The expression of the nucleotide sequence of the 1077 bp fragment of clone 2 by the expression vectors pET28C and pET21C are identified by SEQ ID NO: 135 and SEQ ID NO: 137, respectively.

EXAMPLE 6: EXPRESSION OF CLONE 2 IN *ESCHERICHIA COLI*

The constructs pET28c-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) synthesize, in the bacterial strain BL21 (DE3), a protein fused at the N- and C-terminus for the vector pET28C and the C-terminus for the vector pET21C with 6 Histidines, having an apparent molecular mass of about 45 kDa, identified by SDS-PAGE polyacrylamide gel electrophoresis (SDS = Sodium Dodecyl Sulfate) (Laemmli, 1970 (1)). The reactivity of the protein was demonstrated towards an anti-Histidine monoclonal antibody (DIANOVA) by the Western-blot technique (Towbin et al., 1979 (2)).

The recombinant proteins pET28c-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) were visualized by SDS-PAGE in the insoluble fraction after enzymatic digestion of the bacterial extracts with 50 µl of lysozyme (10 mg/ml) and ultrasound lysis.

The antigenic properties of the recombinant antigens pET28C-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) were tested by Western blotting () [sic] after solubilization of the bacterial pellet with 2% SDS and 50 mM β-mercaptoethanol. After incubation with sera from patients suffering from multiple sclerosis, the sera from neurological controls and the sera from controls at the Blood Transfusion Center (CTS), the immunocomplexes were detected with the aid of an alkaline phosphatase-coupled goat serum anti-human IgG and anti-human IgM.

The results are presented in the table below.

TABLE

Reactivity of sera affected by multiple sclerosis and controls with the MSRV-1 recombinant protein gag clone 2 (1077 bp) = pET21C-clone 2 (1077 bp) and pET28C-clone 2 (1077 bp)^a

DISEASE	NUMBER OF INDIVIDUALS TESTED	NUMBER OF POSITIVE INDIVIDUALS
MS	15	6 2(+++), 2(++), (2(+))
NEUROLOGICAL CONTROLS	2	1(+++)
HEALTHY CONTROLS (CTS)	22	1(+/-)

(a) The strips containing 1.5 µg of recombinant antigen pET-gag clone 2 (1077 bp) exhibit reactivity against sera diluted 1/100. The Western-Blot interpretation is based on the presence or absence of a specific pET-gag clone 2 (1077 bp) band on the strips. Positive and negative controls are included in each experiment.

These results show that, under the technical conditions used, about 40% of the human sera affected by multiple sclerosis which were tested react with the recombinant proteins pET28C-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp). Reactivity was observed on a neurological control and it is of interest to note that the RNAs extracted from this serum, after the reverse transcriptase step, are also amplified by PCR in the pol region. This suggests that people who have not declared MS may also harbor and express this virus. On the other hand, an apparently healthy control (CTS donor) possesses anti-gag (clone 2, 1077 bp) antibodies. This is compatible with an immunity acquired against MSRV-1 independently of a declared associated autoimmune disease.

EXAMPLE 7: PREPARATION OF AN LB13 CLONE CONTAINING IN 3' A PORTION HOMOLOGOUS TO CLONE 2 CORRESPONDING TO THE GAG GENE AND IN 5' A PORTION HOMOLOGOUS TO THE 5M6 CLONE CORRESPONDING TO THE U5 LTR REGION

5 An RT-PCR ("reverse transcriptase-polymerase chain reaction") was carried out using total RNA extracted from virions, obtained from supernatants of B lymphocyte cells of patients suffering from multiple sclerosis, concentrated by ultracentrifugations. The
10 synthesis of cDNA was carried out with a specific primer SEQ No. XXX and the reverse transcriptase "ExpandTM RT" from BOEHRINGER MANNHEIM according to the conditions recommended by the company.

Primer used for the synthesis of the cDNA, identified
15 by SEQ ID NO: 138:

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

A PCR amplification was carried out with Taq polymerase (Perkin Elmer, France) under the following conditions: 94°C 1 min, 55°C 1 min, 72°C 2 min over 35
20 cycles at 72°C for 7 min and with a final reaction volume of 100 µl.

Primers used for the PCR amplification:

- 5' primer, identified by SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

25 - 3' primer, identified by SEQ ID NO: 138

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

A second so-called "seminested" PCR amplification was carried out with a 3' primer situated inside the region already amplified. This second
30 amplification was carried out under the same experimental conditions as those used during the first amplification, using 10 µl of the amplification product derived from the first PCR.

Primers used for the "seminested" PCR amplification:

35 - 5' primer, identified by SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- 3' primer, identified by SEQ ID NO: 140

5' CTA TGT CCT TTT GGA CTG TTT GGG T 3'

The primers SEQ ID NO: 138 and SEQ ID NO: 140 are specific for the gag region, clone 2 nucleotide position No. 373-397 and No. 433-456. The primers used in the 5' region were defined on sequences of the clones obtained during preliminary tests.

An amplification product of 764 bp was obtained and cloned in the following manner:

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kitTM. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCRTM VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning kit[®] (Invitrogen). The mixture was plated after transformation of the ligation into *E. coli* INVαF' bacteria. At the end of the procedure, the white colonies of recombinant bacteria were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "DNA minipreparation" procedure (17). The plasmid preparation of each recombinant colony was cut with the restriction enzyme *EcoRI* and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISMTM Ready Reaction Amplitaq[®] FS, DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The LB13 clone obtained contains an N-terminal region of MSRV-1 gag gene homologous to clone 2 and an

LTR corresponding to a portion of the U5 region. Between the U5 region and gag, a binding site for the transfer RNAs, the PBS "primer binding site", was identified.

5 The nucleotide sequence of the 764 bp fragment of the LB13 clone in the plasmid "pCRTM vector" is represented in the identifier SEQ ID NO: 141.

10 The binding site for the transfer RNAs, having a sequence of PBS tryptophan type, was identified at nucleotide position No. 342-359 of the LB13 clone.

15 As this same PBS was found in the endogenous copies homologous to MSRV1, the endogenous family thus defined is henceforth called HERV W, according to the nomenclature proposed for the endogenous retrovirus families (W=tryptophan).

 A short ORF of about 65 amino acids was found in the U5 region of the 5' LTR of the LB13 clone.

 Sequence of the ORF:

20 PMASNRAITLTAWSKIPFLGIRETKNPRSENTRLATMLEAAHHHFGSSPPLSWEL
 WEQGPQVTIW.

 The corresponding nucleotide sequence starting at an ATG codon is capable of being expressed in a subgenomic DNA from a proviral LTR (U3RU5).

25 Another clone, called LA15, was obtained on the total RNA extracted from virions concentrated by ultracentrifugation from a culture supernatant of synoviocytes obtained from a patient suffering from rheumatoid arthritis. The strategy for amplifying and cloning the LA15 clone is exactly the same which was
30 used for the LB13 clone.

 The nucleotide sequence of the LA15 clone, which is represented in the identifier SEQ ID NO: 142, is very similar to the LB13 clone. This suggests that the MSVR-1 retrovirus detected in multiple sclerosis
35 has sequences which are similar to those found in rheumatoid arthritis.

EXAMPLE 8: RECONSTRUCTION OF AN RU5-GAG REGION FROM THE CLONES LB15, LB13, CL2 AND CL17

The clones CL2 and LB13 have already been described in the preceding examples. The LB15 clone was obtained using the R sequence of the LTR of the cl6 clone in order to define a primer in 5' and the anti-sense primers used are the same as for the LB13 clone. The CL17 clone was obtained by nested RT-PCR using the following primers:

5'-TCATGCAACTGCACTCTTCTGGTCCG-3' (sense)

5'-TCTTGCACTAACCTCCACTGTCCGTTGG-3' (antisense)

5'-ATCCCCCAGTAACAATTTGGTGACCACG-3' (sense)

5'-TCGGGTCTAAGAGGGTACTTCCTTTGGTAGG-3' (antisense)

The LB15 clone was obtained from virions obtained by culturing MS cells. The LB17 clone was obtained from culturing plasma from an MS patient.

These overlapping clones made it possible to reconstruct an RU5-gag sequence with a potential ORF in the gag gene, as presented in Figure 14.

EXAMPLE 9: PREPARATION OF A CLONE 87-23

The region corresponding to integrase was amplified and cloned from MS plasma using a seminested RT PCR with the following primers situated in the pol and env regions of MSRV1.

In the pol region:

5'-TTACGCAGGTCTCAGGGATGAGCTT-3' (sense-primary PCR)

5'-CGGCAGTAGCAGTCTTAGTATCTGAAGCAGTTA-3' (sense-secondary PCR)

In the env region,

5'-GGTACGGAGGGTTTCATGTAGTTTTGAG-3' (anti-sense primary and secondary PCR)

The amplified clone contains 774 bp in the pol/RT region, all the integrase region (1197 bp) and

the start of the env region (480 bp). The nucleotide sequence corresponding to the integrase region and the translation to amino acids of the potential ORF are presented in Figure 15.

5

EXAMPLE 10: CONFIRMATION OF THE PRESENCE OF RNA CONTAINING ENV SEQUENCES RELATED TO ERV9 IN THE RETROVIRAL PARTICLES ASSOCIATED WITH THE MSRV1 GENOME:

Sequences related to ERV9 have been found in a
10 minor proportion in the virion preparations obtained from MS compared with the MSRV1 sequences. The existence of phenomena of co-encapsidation of phylogenetically related endogenous sequences into retroviral particles produced by a replicative strain
15 has been described. Surprisingly, an RNA region comprising an ORF starting in the 3' portion of env and continuing potentially into the 3' LTR has been found in various MS samples. In order to specify the existence of an ORF, transcription-translation tests
20 were carried out and made it possible to show the reality of an env ORF containing the entire transmembrane (TM) portion and ending at the start of the putative LTR. However, an additional frame (ORFX) follows and continues in the 3' LTR. The two products
25 of expression were visualized and their respective ORFs were subcloned. Figure 16 represents the nucleotide and peptide sequences of the B13 clone already described, specifying the ORFs in the truncated env region and in the putative LTR. The presence of such RNAs may be
30 responsible for recombinations with the replicative strain and consequently generate strains having a modified pathogenicity.

BIBLIOGRAPHIC REFERENCES

(1) Laemmli U.K. Cleavage of structural proteins
5 during the assembly of the head of bacteriophage T4.
Nature. (1970). **227**: 680-685.

(2) Towbin H., Staehelin T. & Gordon J.
Electrophoretic transfer of proteins from
10 polyacrylamide gels to nitrocellulose sheets: procedure
and some applications. Proc. Natl. Acad. Sci. USA.
(1979). **76**: 4350-4354.

CLAIMS

1. Nucleic material, in isolated or purified state, comprising a nucleotide sequence chosen from the group which consists of (i) the sequences
5 SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117,
SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130,
SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences
10 equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.
2. Nucleic material, in isolated or purified
15 state, encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118,
20 SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.
3. Retroviral nucleic material, whose pol gene comprises a nucleotide sequence identical or equivalent to a sequence chosen from the group which consists of SEQ ID NO: 112, SEQ ID NO: 124 and their complementary
25 sequences.
4. Retroviral nucleic material, in which the 5' end of the pol gene starts at nucleotide 1419 of SEQ ID NO: 130.
5. Retroviral nucleic material, in which the pol
30 gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with the peptide sequence SEQ ID NO: 113.
6. Retroviral nucleic material, in which the 3'
35 end of the gag gene ends at nucleotide 1418 of SEQ ID NO: 130.
7. Retroviral nucleic material, in which the env gene comprises a nucleotide sequence identical or

equivalent to a sequence chosen from the group which consists of SEQ ID NO: 117, and its complementary sequences.

8. Retroviral nucleic material, in which the env gene comprises a nucleotide sequence which starts at nucleotide 1 of SEQ ID NO: 117 and ends at nucleotide at nucleotide [sic] 233 of SEQ ID NO: 114.

9. Retroviral nucleic material, in which the env gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with the sequence SEQ ID NO: 118.

10. Retroviral nucleic material in which the U3R region of the 3' LTR comprises a nucleotide sequence which ends at nucleotide 617 of SEQ ID NO: 114.

11. Retroviral nucleic material in which the RU5 region of the 5' LTR comprises a nucleotide sequence which starts at nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID NO: 141 or SEQ ID NO: 142.

12. Retroviral nucleic material comprising a sequence which starts at nucleotide 755 of SEQ ID NO: 120 and which ends at nucleotide 617 of SEQ ID NO: 114.

13. Retroviral nucleic material according to any one of the preceding claims, characterized in that it is associated with at least one autoimmune disease such as multiple sclerosis or rheumatoid arthritis.

14. Nucleotide fragment comprising a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.

15. Nucleotide fragment according to Claim 14, consisting of a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, 5 SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous 10 monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.

16. Nucleotide fragment comprising a nucleotide sequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 15 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

17. Nucleotide fragment according to claim 16, 20 consisting of a nucleotide sequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 25 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

18. Nucleic probe for the detection of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, characterized in that it is capable of 30 hybridizing specifically with any fragment according to any one of claims 14 to 17, belonging to the genome of said retrovirus.

19. Probe according to claim 18, characterized in that it possesses from 10 to 100 nucleotides, 35 preferably from 10 to 30 nucleotides.

20. Primer for the amplification, by polymerization, of an RNA or of a DNA of a retrovirus associated with multiple sclerosis and/or rheumatoid

arthritis, characterized in that it comprises a nucleotide sequence identical or equivalent to at least a portion of the nucleotide sequence of a fragment according to any one of claims 8 to 11, in particular a
5 nucleotide sequence having, for every series of 10 contiguous monomers, at least 50%, preferably at least 70% homology with at least said portion of said fragment.

21. Primer according to claim 20, characterized in
10 that its nucleotide sequence is chosen from
SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122,
SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127,
SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, and
SEQ ID NO: 133.

15 22. RNA or DNA, and in particular replication and/or expression vector, comprising a genomic fragment of the nucleic material according to any one of claims 1 to 7 or a fragment according to any one of claims 14 to 17.

20 23. Peptide encoded by any open reading frame belonging to a nucleotide fragment according to any one of claims 14 to 17, in particular a polypeptide, for example oligopeptide forming or comprising an antigenic determinant recognized by sera of patients infected
25 with the MSRV-1 virus, and/or in whom the MSRV-1 virus has been reactivated.

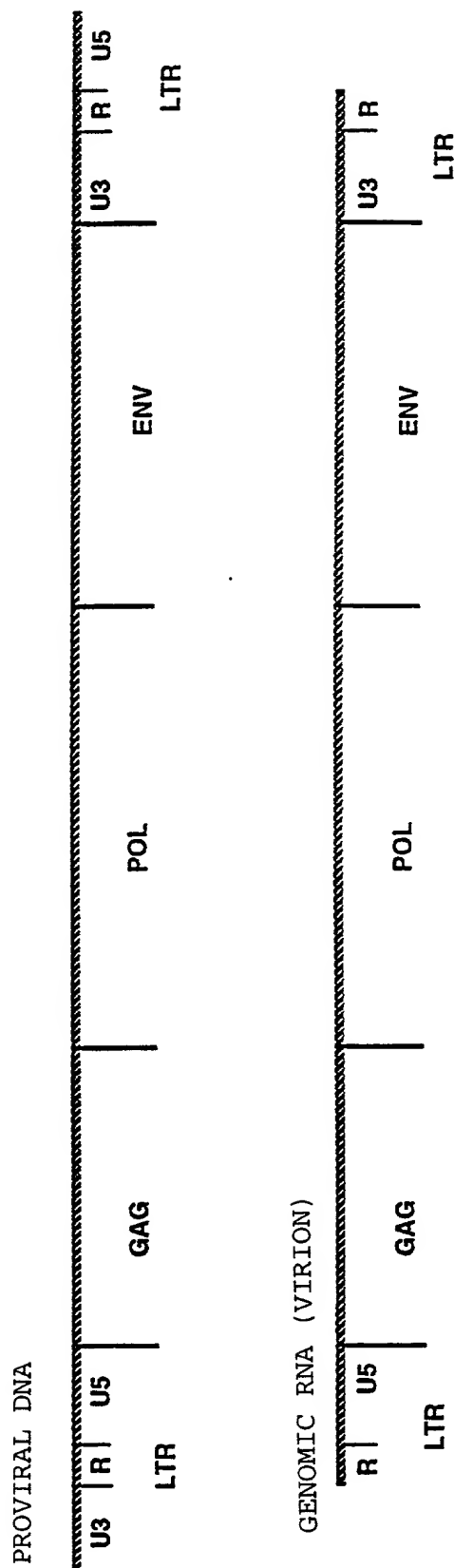
24. Peptide according to claim 23 comprising a sequence identical, partially or completely, or equivalent to a sequence chosen from SEQ ID NO: 113,
30 SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

25. Diagnostic, prophylactic or therapeutic composition, in particular for inhibiting the expression of at least one retrovirus associated with
35 multiple sclerosis and/or rheumatoid arthritis, comprising a nucleotide fragment according to any one of claims 14 to 17.

26. Method for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, characterized in that an RNA and/or a DNA assumed to belong to or obtained from said
- 5 retrovirus, or their complementary RNA and/or DNA, is brought into contact with a composition comprising a nucleotide fragment according to any one of claims 14 to 17.

1 / 32

FIG 1



2 / 32
FIG 2

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCTTATAGAA	GGACCCCTAG	TATGGGGTAA	TCCCTCTCGG	GAAACCAAGC	50
A Y R R	T P S M G .	S P L G	N Q A		
L I E	G P L V	W G N	P L W	E T K P	
L .	K D P .	Y G V I	P S G	K P S	
CCAGTACTC	AGCAGGAAAA	ATAGAATAGG	AAACCTCACA	AGGACATACT	100
P V L	S R K N	R I G	N L T	R T Y F	
Q Y S	A G K I E .	E T S Q	G H T		
P S T Q	Q E K .	N R	K P H K	D I L	
TTCCTCCCCCT	CCAGATGGCT	AGCCACTGAG	GAAGGAAAAA	TACTTTCACC	150
P P L	Q M A S H .	G R K N	T F T		
F L P S	R W L A T E	E G K I	L S P		
S S P	P D G .	P L R	K E K	Y F H L	
TGCAGCTAAC	CAACAGAAAT	TACTTAAAC	CCTTCACCAA	ACCTTCCACT	200
C S .	P T E I T .	N P S P N	L P L		
A A N	Q Q K L	L K T	L H Q	T F H L	
Q L T	N R N	Y L K P	F T K	P S T	
TAGGCATTGA	TAGCACCCAT	CAGATGGCCA	AATTATTATT	TACTGGACCA	250
R H .	. H P S	D G Q	I I I	Y W T R	
G I D	S T H	Q M A K	L L F	T G P	
. A L I	A P I	R W P	N Y Y L	L D Q	
GGCTTTTICA	AAACTATCAA	GAAGATAGTC	AGGGGCTGTG	AAGTGTGCCA	300
P F Q	N Y Q	E D S Q	G L .	S V P	
G L F K	T I K	K I V	R G C E	V C Q	
A F S	K L S R	R .	S G A V	K C A K	
AAGAAATAAT					310
K K .					
R N N					
E I					

3 / 32
FIG 2 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCCTGATCT	TTAACCTCCT	TGTTAAGTIT	GTCCTCTCCA	GAATCAAAAC	50
P C I F	N L L	V K F	V S S R	I K T	
P V S	L T S L	L S L	S L P	E S K L	
L Y L	. P P	C . V C	L F Q	N Q N	
TGTAATACTA	CAAATTGTC	TTCAAATGGA	GCACCAGATG	GAGTCCATGA	100
V K L	Q I V L	Q M E	H Q M	E S M T	
. N Y	K L F	F K W S	T R W	S P .	
C K T T	N C S	S N G	A P D G	V H D	
CTAAGATCCA	CCGTGGACCC	CTGGACCGGC	CTGCTAGCCC	ATGCTCCGAT	150
K I H	R G P	L D R P	A S P	C S D	
L R S T	V D P	W T G	L L A H	A P M	
. D P	P W T P	G P A C	. P	M L R C	
GTTAATGACA	TTGAAGGCAC	CCCTCCCGAG	GAAATCTCAA	CTGCACAACC	200
V N D I	E G T	P P E	E I S T	A Q P	
L M T	L K A P	L P R	K S Q	L H N P	
. . H	. R H	P S R G	N L N	C T T	
CCTACTATGC	CCCAATTTCAG	CGGGAAGCAG	TTAGAGCGGT	CATCAGCCAA	250
L L C	P N S A	G S S	. S G	H Q P T	
Y Y A	P I Q	R E A V	R A V	I S Q	
P T M P	Q F S	G K Q	L E R S	S A N	
CCTCCCCAAC	AGCACTTGGG	TTTTCCTGTT	GAGAGGGGGG	ACTGAGAGAC	300
S P T	A L G	F S C	. E G G	L R D	
P P Q Q	H L G	F P V	E R G D	. E T	
L P N	S T W V	F L L	R G G	T E R Q	
AGGACTAGCT	GGATTTCCTA	GGCCAACGAA	GAATCCCTAA	GCCTAGCTGG	350
R T S W	I S .	A N E	E S L S	L A G	
G L A	G F P R	P T K	N P .	A . L G	
D . L	D F L	G Q R R	I P K	P S W	

4 / 32
FIG 3

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GAAGGTGACT	GCATCCACCT	CTAAACATGG	GGCTTGCAAC	TTAGCTCACA	400
K V T	A S T S	K H G	A C N	L A H T	
R . L	H P P	L N M G	L A T	. L T	
E G D C	I H L	. T W	G L Q L	S S H	
CCCCACCAAT	CAGAGAGCTC	ACTAAAATGC	TAATTAGGCA	AAAATAGGAG	450
R P I	R E L	T K M L	I R Q K	. E	
P D Q S	E S S	L K C	. L G K	N R R	
P T N	Q R A H	. N A N	. A K I G G		
GTAAAGAAAT	AGCCAATCAT	CTATTGCCTG	AGAGCACAGC	GGGAGGGACA	500
V K K .	P I I	Y C L	R A Q R	E G Q	
. R N	S Q S S	I A .	E H S	G R D K	
K E I	A N H	L L P E	S T A	G G T	
AGGATCGGGA	TATAAACCCA	GGCATTGAG	COGGCAACGG	CAACCCCCCTT	550
G S G	Y K P R	H S S	R Q R	Q P P L	
D R D	I N P	G I R A	G N G	N P L	
R I G I	. T Q	A F E	P A T A	T P F	
TGGGTCOOCT	COCTTGTAT	GGGCGCTCTG	TTTTCACCTCT	ATTTCACCTCT	600
G P L	P L Y	G R S V	F T L	F H S	
W V P S	L C M	G A L	F S L Y	F T L	
G S P	P F V W	A L C	F H S	I S L Y	
ATTAAATCTT	GCAACTGAAA	AAAAAAAAAA	AAAAA		635
I K S C	N . K	K K K K	K		
L N L	A T E K	K K K K	K		
. I L	Q L K	K K K K	K		

5 / 32
FIG 4

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCCCTCC	CTTATCATA	TTTCTCTTT	ACIGTCTCT	TACCCCTTT	50
M A L P	Y H T	F L F	T V L L	P P F	
W P S	L I I L	F S L	L F S	Y P L S	
G P P	L S Y	F S L Y	C S L	T P F	
CGCTCTCACT	GCACCCCTC	CATGCTGCTG	TACAACCAGT	AGCTCCCTT	100
A L T	A P P P	C C C	T T S	S S P Y	
L S L	H P L	H A A V	Q P V	A P L	
R S H C	T P S	M L L	Y N Q	L P L	
ACCAAGAGIT	TCTATGAAG	ACGCGCTTC	CTGGAAATAT	TGATGCCCCA	150
Q E F L	R T R L P	G N I	D A P		
T K S F	Y E E	R G F	L E I L	M P H	
P R V	S M K N	A A S	W K Y	C P I	
TCATATAGGA	GTTTATCTAA	GGGAAACTCC	ACCTTCACTG	CCCACACCCA	200
S Y R S	L S K	G N S	T F T A	H T H	
H I G	V Y L R	E T P	P S L	P T P I	
I . E	F I .	G K L H	L H C	P H P	
TATGCCCCGC	AACCTGCTATA	ACTCTGCCAC	TCTTTGCATG	CATGCAAATA	250
M P R	N C Y N	S A T	L C M	H A N T	
C P A	T A I	T L P L	F A C	M Q I	
Y A P Q	L L .	L C H	S L H A	C K Y	
CTCATTATTG	GACAGGGAAA	ATGATTAATC	CTAGTTGTCC	TGGAGGACTT	300
H Y W	T G K	M I N P	S C P	G G L	
L I I G	Q G K	L I	L V V L	E D L	
S L L	D R E N	D . S	L S	W R T W	
GGAGCCACTG	TCTGTGGAC	TTACTTCACC	CATACCAGTA	TGCTGTATGG	350
G A T V	C W T	Y F T	H T S M	S D G	
E P L	S V G L	T S P	I P V	C L M G	
S H C	L L D	L L H P	Y Q Y	V . W	

6 / 32
FIG 4 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ACCTCACCTIG	TGTAAAATTT	AGCAATACTA	TAGACACAAC	CAGCTCCCAA	750
L T C	V K F	S N T I	D T T	S S Q	
T S P V	. N L	A I L	. T Q P	A P N	
P H L	C K I .	Q Y Y	R H N	Q L P M	
TGCATCAGGT	GGGTAAACACC	TCCACACCA	ATAGTCTGCC	TACCCTCAGG	800
C I R W	V T P	P T R	I V C L	P S G	
A S G	G . H L	P H E	. S A	Y P Q E	
H Q V	G N T	S H T N	S L P	T L R	
AATATTTTTT	GTCTGIGGTA	CCTCAGCCTA	TCATTGTTTG	AATGGCTCTT	850
I F F	V C G T	S A Y	H C L	N G S S	
Y F L	S V V	P Q P I	I V .	M A L	
N I F C	L W Y	L S L	S L F E	W L F	
CAGAATCTAT	GTCCTTCCTC	TCATTCTTAG	TGCCCCCTAT	GACCATCTAC	900
E S M	C F L	S F L V	P P M	T I Y	
Q N L C	A S S	H S .	C P L .	P S T	
R I Y	V L P L	I L S	A P Y	D H L H	
ACTGAACAAG	ATTTATACAA	TCATGTGGTA	CCTAAGCCCC	ACAACAAAAG	950
T E Q D	L Y N	H V V	P K P H	N K R	
L N K	I Y T I	M S Y	L S P	T T K E	
. T R	F I Q	S C R T	. A P	Q Q K	
AGTACCCATT	CTTCCTTTTG	TTATCAGAGC	AGGAGTGCTA	GGCAGACTAG	1000
V P I	L P F V	I R A	G V L	G R L G	
Y P F	F L L	L S E Q	E C .	A D .	
S T H S	S F C	Y Q S	R S A R	Q T R	
GTACTGGCAT	TGGCAGTATC	ACAACCTCTA	CTCAGTTCTA	CTACAAACTA	1050
T G I	G S I	T T S T	Q F Y	Y K L	
V L A L	A V S	Q P L	L S S T	T N Y	
Y W H	W Q Y H	N L Y	S V L	L Q T I	

FIG 4 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCTCAAGAAA	TAAATGGTGA	CATGGAACAG	GTCACITGACT	CCCTGGTTCAC	1100
S Q E I	N G D	M E Q	V T D S	L V T	
L K K	. M V T	W N R	S L T	P W S P	
S R N	K W .	H G T G	H . L	P G H	
CTTGCAAGAT	CAACTTAACT	CCCTAGCAGC	AGTAGTCCTT	CAAAATCGAA	1150
L Q D	Q L N S	L A A	V V L	Q N R R	
C K I	N L T	P . Q Q	. S F	K I E	
L A R S	T . L	P S S	S S P S	K S K	
GAGCTTTAGA	CTTGCTAACC	GCCAAAAGAG	GGGGAACCTG	TTTATTTTAA	1200
A L D	L L T	A K R G	G T C	L F L	
E L .	T C .	P P K E	G . E P V	Y F .	
S F R	L A N R	Q K R	G N L	F I F R	
GGAGAAGAAC	GCTGTATTAA	TGTTAATCAA	TCCAGAATTG	TCACTGAGAA	1250
G E E R	C Y Y	V N Q	S R I V	T E K	
E K N	A V I M	L I N	P E L	S L R K	
R R T	L L L	C . S I	Q N C	H . E	
AGTTAAAGAA	ATTGAGATC	GAATACAATG	TAGAGCAGAG	GAGCTTCAAA	1300
V K E	I R D R	I Q C	R A E	E L Q N	
L K K	F E I	E Y N V	E Q R	S F K	
S . R N	S R S	N T M	. S R G	A S K	
ACACCGAAG	CTGGGGCCTC	CTCAGCCAAT	GGATGCCCTG	GGTTCCTCCC	1350
T E R	W G L	L S Q W	M P W	V L P	
T P N A	G A S	S A N	G C P G	F S P	
H R T	L G P P	Q P M	D A L	G S P L	
TTCTTAGGAC	CTCTAGCAGC	TCTAATATTG	TIACTCCTCT	TTGGACCTG	1400
F L G P	L A A	L I L	L L L F	G P C	
S . D	L . Q L	. Y C	Y S S	L D P V	
L R T	S S S	S N I V	T P L	W T L	

8 / 32

FIG 4 (continued)

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
TATCTTTAAC	CTCCTTGTTA	AGTTTGCTC	TTCCAGAATT	GAAGCTGTAA
I F N	L L V K	F V S	S R I	E A V K
S L T	S L L	S L S L	P E L	K L .
Y L .	P P C .	V C L	F Q N .	S C K

AGCTACAGAT	GGTCTTACAA	ATGGAACCCC	A	1481
L Q M	V L Q	M E P		
S Y R W	S Y K	W N P		
A T D	G L T N	G T P		

9 / 32
FIG 5

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCAAAATCGA	AGAGCTTTAG	ACTTGCTAAC	CGCCAAAAGA	GGGGGAACCT	50
S K S K	S F R	L A N	R Q K R	G N L	
Q N R	R A L D	L L T	A K R	G G T C	
K I E	E L .	T C .	P P K E	G E P	
GTTTATTTTT	AGGGGAAGAA	TGCTGTTAGT	ATGTTAATCA	ATCTGGAATC	100
F I F	R G R M	L L V	C . S	I W N H	
L F L	G E E	C C .	Y V N Q	S G I	
V Y F .	G K N	A V S	M L I N	L E S	
ATTACTGAGA	AAGTTAAAGA	AATTTGAGAT	CGAATATAAT	GTAGAGCAGA	150
Y . E	S . R	N L R S	N I M .	S R	
I T E K	V K E	I . D	R I . C	R A E	
L L R	K L K K	F E I	E Y N	V E Q R	
GGACCTTCAA	AACACTGCAC	CCTGGGGGCT	CCTCAGCCAA	TGGATGCGCT	200
G P S K	H C T	L G P	P Q P M	D A L	
D L Q	N T A P	W G L	L S Q	W M P W	
T F K	T L H	P G A S	S A N	G C P	
GGACTCTCCC	CTTCTTAGGA	CCTCTAGCAG	CTATAATATT	TTTACTCCTC	250
D S P	L L R T	S S S	Y N I	F T P L	
T L P	F L G	P L A A	I I F	L L L	
G L S P	S . D	L . Q	L . Y F	Y S S	
TTTGGACCGT	GTATCTTCAA	CTTCTTGTT	AAGTTTGICT	CTTCCAGAAT	300
W T L	Y L Q	L P C .	V C L	F Q N	
F G P C	I F N	F L V	K F V S	S R I	
L D P	V S S T	S L L	S L S	L P E L	
TGAAGCTGTA	AAGCTACAAA	TAGTCTTCA	AATGGAACCC	CAGATGCAGT	350
. S C K	A T N	S S S	N G T P	D A V	
E A V	K L Q I	V L Q	M E P	Q M Q S	
K L .	S Y K .	F F K	W N P	R C S	

10 / 32
FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCATGACTAA	AATCTACCGT	GGACCCCTGG	ACCGGCCTGC	TAGACTATGC	400
H D .	N L P W	T P G	P A C	. T M L	
M T K	I Y R	G P L D	R P A	R L C	
P . L K	S T V	D P W	T G L L	D Y A	
TCTGATGTTA	ATGACATTGA	AGTCACCCCT	CCCGAGGAAA	TCTCAACTGC	450
. C .	. H .	S H P S	R G N	L N C	
S D V N	D I E	V T P	P E E I	S T A	
L M L	M T L K	S P L	P R K	S Q L H	
ACAACCCCTA	CTACACTCCA	ATTCACTAGG	AAGCAGTTAG	AGCAGTTGTC	500
T T P T	T L Q	F S R	K Q L E	Q L S	
Q P L	L H S N	S V G	S S .	S S C Q	
N P Y	Y T P	I Q .	E A V R	A V V	
AGCCAACCTC	CCCAACAGTA	CTTGGGTTTT	CCTGTGAGCA	GGGTGGACTG	550
A N L	P N S T	W V F	L L R	G W T E	
P T S	P T V	L G F S	C . E	G G L	
S Q P P	Q Q Y	L G F	P V E R	V D .	
AGAGACAGGA	CTAGCTGGAT	TTCCTAGGCT	GACTAAGAAT	CCCAAGCCT	600
R Q D	. L D	F L G .	L R I	P K P	
R D R T	S W I	S . A	D . E	S X S L	
E T G	L A G F	P R L	T K N	P X A X	
ANCTGGGAAG	GTCACCGCAT	CCATCTTTAA	ACATGGGGCT	TGCAACTTAG	650
X W E G	D R I	H L .	T W G L	Q L S	
X G K	V T A S	I F K	H G A	C N L A	
L G R	. P H	P S L N	M G L	A T .	
CTCACACCCG	ACCAATCAGA	GAGCTCACTA	AAATGCTAAT	CAGGCAAAAA	700
S H P	T N Q R	A H .	N A N	Q A K T	
H T R	P I R	E L T K	M L I	R Q K	
L T P D	Q S E	S S L	K C .	S G K N	

11 / 32
FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAGGAGGTAA	AGCAATAGCC	AATCATCTAT	TGCTGAGAG	CACAGCGGGA	750
G G K	A I A	N H L L	P E S	T A G	
Q E V K	Q . P	I I Y	C L R A	Q R E	
R R .	S N S Q	S S I A	. E H	S G K	
AGGACAAGGA	TTGGGATATA	AACTCAGGCA	TTCAAGCCAG	CAACAGCAAC	800
R T R I	G I .	T Q A	F K P A	T A T	
G Q G	L G Y K	L R H	S S Q	Q Q Q P	
D K D	W D I	N S G I	Q A S	N S N	
CCCCCTTGGG	TCCCCCTCCA	TTGTATGGGA	GCTCTGTTTT	CACTCTATTT	850
P F G	S P P I	V W E	L C F	H S I S	
P L G	P L P	L Y G S	S V F	T L F	
P L W V	P S H	C M G	A L F S	L Y F	
CACTCTATTA	AATCATGCAA	CTGCACTCTT	CTGGTCCGIG	TTTTTTATGG	900
L Y .	I M Q	L H S S	G P C	F L W	
H S I K	S C N	C T L	L V R V	F Y G	
T L L	N H A T	A L F	W S V	F F M A	
CTCAAGCTGA	GCTTTTGTTC	GCCATCCACC	ACTGCTGITT	GCCACCGTCA	950
L K L S	F C S	P S T	T A V C	H R H	
S S .	A F V R	H P P	L L F	A T V T	
Q A E	L L F	A I H H	C C L	P P S	
CAGACCCGCT	GCTGACTTCC	ATCCCTTGGG	ATCCAGCAGA	GIGTCCACTG	1000
R P A	A D F H	P F G	S S R	V S T V	
D P L	L T S	I P L D	P A E	C P L	
Q T R C	. L P	S L W	I Q Q S	V H C	
TGCTCTGAT	CCAGCGAGGT	ACCCATTGCC	ACTCCCGATC	AGGCTAAAGG	1050
L L I	Q R G	T H C H	S R S	G . R	
C S .	S S E V	P I A	T P D Q	A K G	
A P D	P A R Y	P L P	L P I	R L K A	

12 / 32
FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CTTGCCATTG	TTCCTGCGATG	GCTAAGTGGC	TGGGTTTGTG	CTAATAGAAC	1100
L A I V	P A W	L S A	W V C P	N R T	
L P L	F L H G	. V P	G F V	L I E L	
C H C	S C M	A K C L	G L S	. . N	
TGAACACTGG	TCACTGGGTT	CCATGGTTCT	CTTCCATGAC	CCACGGCTTC	1150
E H W	S L G S	M V L	F H D	P R L L	
N T G	H W V	P W F S	S M T	H G F	
. T L V	T G F	H G S	L P .	P T A S	
TAATAGAGCT	ATAACACTCA	CCGCATGGCC	CAAGATTCCA	TTCCTTGGTA	1200
I E L	. H S	P H G P	R F H	S L V	
. . S Y	N T H	R M A	Q D S I	P W Y	
N R A	I T L T	A W P	K I P	F L G I	
TCTGTGAGGC	CAAGAACCCC	AGGTCAGAGA	ANGTGAGGCT	TGCCACCATT	1250
S V R P	R T P	G Q R	X . G L	P P F	
L . G	Q E P Q	V R E	X E A	C H H L	
C E A	K N P	R S E X	V R L	A T I	
TGGGAAGTGG	CCCACTGCCA	TTTTGGTACG	GGCCCCACCAC	CATCTTGGGA	1300
G K W	P T A I	L V A	A H H	H L G S	
G S G	P L P	F W .	R P T T	I L G	
W E V A	H C H	F G S	G P P P	S W E	
GCTGTGGGAG	CAAGGATCCC	CCAGTAACA			1329
C G S	K D P	P V T			
A V G A	R I P	Q .			
L W E	Q G S P	S N			

13 / 32

FIG 6

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
OCTAGAAGCT	ATTCTGGAGA	ATTGGGACCA	ATGTCACACT	CAGACGCTAA	50
P R T Y	S G E	L G P	M . H S	D A K	
L E R	I L E N	W D Q	C D T	Q T L R	
. N V	F W R	I G T N	V T L	R R .	
GAAAGAAAG	ATTTATATTC	TTCTGCAGTA	CCGCTGGCC	ACAATATCCT	100
K E T	I Y I L	L Q Y	R L A	T I S S	
K K R	F I F	F C S T	A W P	Q Y P	
E R N D	L Y S	S A V	P P G H	N I L	
CTTCAAGGA	GAGAAACCTG	GCTTCTGAG	GGAAGTATAA	ATTATAACAT	150
S R E	R N L	A S . G	K Y K	L . H	
L Q G R	E T W	L P E	G S I N	Y N I	
F K G	E K P G	F L R	E V .	I I T S	
CATCTTACAG	CTAGACCTCT	TCTGTAGAAA	GGAGGGCAAA	TGGAGTGAAG	200
H L T A	R P L	L . K	G G Q M	E . S	
I L Q	L D L F	C R K	E G K	W S E V	
S Y S	. T S	S V E R	R A N	G V K	
TGCCATATGT	GCAAACTTTC	TTTTCATTAA	GAGACAATC	ACAATTATGT	250
A I C	A N F L	F I K	R Q L	T I M .	
P Y V	Q T F	F S L R	D N S	Q L C	
C H M C	K L S	F H .	E T T H	N Y V	
AAAAAGTGTG	GTTTATGCCC	TACAGGAAGC	CCTCAGAGTC	CACCTCCCTA	300
K V W	F M P	Y R K P	S E S	T S L	
K K C G	L C P	T G S	P Q S P	P P Y	
K S V	V Y A L	Q E A	L R V	H L P T	
CCCCAGGTC	CCCTCCCCGA	CTCTTCTCTC	AACTAATAAG	GACCCCCCTT	350
P Q R P	L P D	S F L	N . . G	P P F	
P S V	P S P T	P S S	T N K	D P P L	
P A S	P P R	L L P Q	L I R	T P L	
TAAOCCAAAC	GGTCCAAAAG	GAGATAGACA	AAGGGGTAAA	CAATGAACCA	400
N P N	G P K G	D R Q	R G K	Q . T K	
T Q T	V Q K	E I D K	G V N	N E P	
. P K R	S K R	R . T	K G .	T M N Q	
AAGAGTCCA	ATATTCCCCG	ATTATGCCCC	CTCCAAGCAG	TGAGAGGAG	450
E C Q	Y S P	I M P P	P S S	E R R	
K S A N	I P R	L C P	L Q A V	R G G	
R V P	I F P D	Y A P	S K Q	. E E E	
AGAATTGGC	CCAGCCAGAG	TGCTGTACC	TTTTTCTCTC	TCAGACTTAA	500
R I R P	S Q S	A C T	F F S L	R L K	
E F G	P A R V	P V P	F S L	S D L K	
N S A	Q P E	C L Y L	F L S	Q T .	

14 / 32
FIG 6 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGCAAATTAA	AATAGACCTA	GGTAAATTCT	CAGATAACCC	TGACGGCTAT	550
A N .	N R P R .	I L R .	P .	R L Y	
Q I K	I D L	G K F S	D N P	D G Y	
S K L K .	T .	V N S	Q I T L	T A I	
ATTGATGTTT	TACAAGGGTT	AGGACAATCC	TTTGATCTGA	CATGGAGAGA	600
. C F	T R V	R T I L .	S D	M E R	
I D V L	Q G L	G Q S	F D L T	W R D	
L M F	Y K G .	D N P	L I .	H G E I	
TATAATGTTA	CTACTAAATC	AGACACTAAC	CCCAAATGAG	AGAAGTGCCG	650
Y N V T	T K S	D T N	P K .	E K C R	
I M L	L L N Q	T L T	P N E	R S A A	
. C Y Y .	I R H .	P	Q M R	E V P	
CTGTAACTGC	AGCCCCGAGAG	TTTGGCGATC	TTTGGTATCT	CAGTCAGGCC	700
C N C	S P R V	W R S	L V S	Q S G Q	
V T A	A R E	F G D L	W Y L	S Q A	
L .	L Q	P E S	L A I	F G I S	V R P
AACAATAGGA	TGACAACAGA	GGAAAGAACA	ACTCCACAG	GCCAGCAGGC	750
Q .	D D N R	G K N N	S H R	P A G	
N N R M	T T E	E R T	T P T G	Q Q A	
T I G .	Q Q R	K E Q	L P Q	A S R Q	
AGTTCOCAGT	GTAGACCCCTC	ATTGGGACAC	AGAATCAGAA	CATGGAGATT	800
S S Q C	R P S	L G H	R I R T	W R L	
V P S	V D P H	W D T	E S E	H G D W	
F P V .	T L	I G T Q	N Q N	M E I	
GGTGCCACAA	ACATTGCTA	ACTTGGGTGC	TAGAAGCACT	GAGGAAACT	850
V P Q	T F A N	L R A	R R T	E E N .	
C H K	H L L	T C V L	E G L	R K T	
G A T N	I C .	L A C .	K D .	G K L	
AGGAGAAGC	CTATGAATTA	CICAATGATG	TCCACTATAA	CACAGGAAA	900
E E A	Y E L	L N D V	H Y N	T G K	
R K K P	M N Y	S M M	S T I T	Q G K	
G R S	L .	I T Q .	C P L .	H R E R	
GGAAGAAAT	CTTACTGCTT	TTCTGGACAG	ACTAAGGAG	GCAITGAGGA	950
G R K S	Y C F	S G Q	T K G G	I E E	
E E N	L T A F	L D R	L R E	A L R K	
K K I	L L L	F W T D .	G R H .	G	
AGCATACCTC	CCGTGTCACCT	GACTCTATTG	AAGGCCAACT	AATCTTAAAG	1000
A Y L	P V T .	L Y .	R P T	N L K G	
H T S	L S P	D S I E	G Q L	I L K	
S I P P	C H L	T L L	K A N .	S . R	

15 / 32
FIG 6 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GATAAGTTTA	TCATCAGTC	AGCTGCAGAC	ATTAGAAAAA	ACTTCAAAAG	1050
. V Y H S V	S C R H .	K K L Q K			
D K F I T Q S	A A D I R K N	F K S			
I S L S L S Q	L Q T L E K	T S K V			
TCTGCTTAG	GCGCGAGCA	GAACCTAGAA	ACCTTATTTA	ACTTGGCATC	1100
S A L G P E Q	N L E T L F N	L A S			
L P . A R S R	T . K P Y L	T W H P			
C L R P G A	E L R N P I .	L G I			
CTCAGTTTTT	TATAATAGAG	ATCAGGAGGA	GCAGGCGAAA	CGGGACAAAC	1150
S V F Y N R D	Q E E Q A K	R D K R			
Q F F I I E	I R R S R R N	G T N			
L S F L . .	R S G G A G E T	G Q T			
GGGATAAAAA	AAAAAGGGGG	GGTCCACTAC	TTTAGTCATG	GCCCTCAGGC	1200
D K K K R G	G P L L .	S W P S G			
G I K K K G G	V H Y F S H .	G P Q A			
G . K K K G G	S T T L V M	A L R Q			
AAGCAGACTT	TGGAGGCTCT	GCAAAAGGGA	AAAGCTGGGC	AAATCAAATG	1250
K Q T L E A L	Q K G K A G Q	I K C			
S R L W R L C	K R E K L G	K S N A			
A D F G G S	A K G K S W A	N Q M			
OCTAATAGGG	CTGGCTTCCA	GTCGGGTCTA	CAAGGACACT	TTAAAAAGA	1300
L I G L A S S	A V Y K D T	L K K I			
. . G W L P	V R S T R T L	. K R			
P N R A G F Q	C G L Q G H F	K K D			
TTATCCAAGT	AGAATAAGC	CGCCCCCTTG	TCCATGCCCC	TTACGTCAAG	1350
I Q V E I S	R P L V H A P	Y V K			
L S K . K .	A A P L S M P L	T S R			
Y P S R N K P	P P C P C P	L R Q G			
GGAATCACTG	GAAGGCCCC	TGCCCCAGGG	GATGAAGATA	CCTCAGTCA	1400
G I T G R P T	A P G D E D T	L S Q			
E S L E G P L	P Q G M K I	L . V R			
N H W K A H	C P R G .	R Y S E S			
GAAGCATTA	ACCAGATGAT	CCAGCAGCAG	GACTGAGGGT	GCCCCGGGG	1450
K P L T R .	S S S R T E G	A R G E			
S H . P D D	P A A G L R V	P G A			
E A I N Q M I	Q Q Q D .	G C P G R			
AGGGCCAGCC	CATGCCATCA	CCCTCAGAGA	GCCCCGGGTA	TGTTTGAACA	1500
R Q P M P S	P S Q S P G Y	V . P			
S A S P C H H	P H R A P G M	F D H			
A P A H A I T	L T E P R V	C L T I			

16 / 32
FIG 6 (continued)

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
TTGAGAGCCA A				1511
L R A				
. E P				
E S Q				

17 / 32

FIG 7

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGGCAGCA	GCCATCATCA	TCATCATCAC	AGCAGCGGCC	TGGTGGCGCG	50
M G S S	H H H	H H H	S S G L	V P R	
CGGCAGGCAT	ATGGCTAGCA	TGACTGGTGG	ACAGCAAATG	GGTCGGATCC	100
G S H	M A S M	T G G	Q Q M	G R I L	
TAGAAOGTAT	TCTGGAGAAT	TGGGACCAAT	GTCACACTCA	GACGCTAAGA	150
E R I	L E N	W D Q	C D T	Q T L R	
AAGAAACGAT	TTATATTCTT	CTGCAGTACC	GCCTGGCCAC	AATATCCTCT	200
K K R F	I F F	C S T	A W P Q	Y P L	
TCAAGGGAGA	GAAACCTGGC	TTCCTGAGGG	AAGTATAAAT	TATAACATCA	250
Q G R	E T W L	P E G	S I N	Y N I I	
TCTTACAGCT	AGACCTCTTC	TGTAGAAAGG	AGGGCAAATG	GAGTGAAGTG	300
L Q L	D L F	C R K E	G K W	S E V	
CCATATGIGC	AAACTTTCTT	TTCATTGAAG	GACAACCTAC	AATTATGTAA	350
P Y V Q	T F F	S L R	D N S Q	L C K	
AAGTGTGGT	TTATGCCCCA	CAGGAAGCCC	TCAGAGTCCA	CCTCCCCTACC	400
K C G	L C P T	G S P	Q S P	P P Y P	
CCAGGGTCCC	CTCCCCGACT	CCTTCTCTCA	CTAATAAGGA	CCCCCTTTTA	450
S V P	S P T	P S S T	N K D	P P L	
ACCCAAACGG	TCCAAAAGGA	GATAGACAAA	GGGGTAAACA	ATGAACCAAA	500
T Q T V	Q K E	I D K	G V N N	E P K	
GAGTGCATAT	ATTCCCCGAT	TATGCCCCCT	CCAAGCAGTG	AGAGGAGGAG	550
S A N	I P R L	C P L	Q A V	R G G E	
AATTGGGCCC	AGCCAGAGTG	CCGTGACCTT	TTTCTCTCTC	AGACTTAAAG	600
F G P	A R V	P V P F	S L S	D L K	
CAAATTAATA	TAGACCTAGG	TAAATTCTCA	GATAACCTTG	ACGGCTATAT	650
Q I K I	D L G	K F S	D N P D	G Y I	
TGATGTTTAA	CAAGGGTTAG	GACAATCCTT	TGATCTGACA	TGGAGAGATA	700
D V L	Q G L G	Q S F	D L T	W R D I	
TAATGTTACT	ACTAAATCAG	ACACTAAACC	CAAATGAGAG	AAGTGGCGCT	750
M L L	L N Q	T L T P	N E R	S A A	
GTAAGTGCAG	CCCGAGAGTT	TGGGATCTTT	TGGTATCTCA	GTCAGGOCAA	800
V T A A	R E F	G D L	W Y L S	Q A N	

18 / 32
FIG 7 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAATAGGATG	ACAACAGAGG	AAAGAACAAC	TCCCACAGGC	CAGCAGGCAG	850
N R M	T T E E	R T T	P T G	Q Q A V	
TTCCCAGTGT	AGACCCTCAT	TGGGACACAG	AATCAGAACA	TGGAGATTGG	900
P S V	D P H	W D T E	S E H	G D W	
TGCCACAAAC	ATTTTGCTAAC	TTGGGTGCTA	GAAGGACTGA	GGAAACTAG	950
C H K H	L L T	C V L	E G L R	K T R	
GAAGAAGCCT	ATGAATTACT	CAATGATGTC	CACTATAACA	CAGGGAAAGG	1000
K K P	M N Y S	M M S	T I T	Q G K E	
AAGAAAATCT	TACTGCTTTT	CTGGACAGAC	TAAGGGAGGC	ATTGAGGAAG	1050
E N L	T A F	L D R L	R E A	L R K	
CATACCTCCC	TGTCACCTGA	CCTATTTGAA	GGCCAACTAA	TCTTAAAGGA	1100
H T S L	S P D	S I E	G Q L I	L K D	
TAAGTTTATC	ACTCAGTCAG	CTGCAGACAT	TAGAAAAAAC	TTCAAAAGTC	1150
K F I	T Q S A	A D I	R K N	F K S L	
TGCTTAAGCT	TGCGGGCCCA	CTCGAGCACC	ACCACCACCA	CCACTGAGAT	1200
P K L	A A A	L E H H	H H H	H . D	
CCGGCTTGCTA	ACAAAGCCCG	AAAGGAAGCT	GAGTTGGCTN	GTGGCNA	1247
P A A N	K A R	K E A	E L A X	G	

19 / 32
FIG 8

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCTAGCA	TGACTGGTGG	ACAGCAAATG	GGTCGGATCC	TAGAACGTAT	50
M A S M	T G G	Q Q M	G R I L	E R I	
TCTGGAGAAT	TGGGACCAAT	GTGACACTCA	GACGCTAAGA	AAGAAACGAT	100
L E N	W D Q C	D T Q	T L R	K K R F	
TTATATTCIT	CTGCAGTACC	GCCTGGCCAC	AATATCCTCT	TCAAGGGAGA	150
I F F	C S T	A W P Q	Y P L	Q G R	
GAAACCTGGC	TTCTTGAGGG	AAGTATAAAT	TATAACATCA	TCTTACAGCT	200
E T W L	P E G	S I N	Y N I I	L Q L	
AGACCTCTTC	TGTAGAAAGG	AGGGCAAATG	GAGTGAAGTG	CCATATGTGC	250
D L F	C R K E	G K W	S E V	P Y V Q	
AAACTTTCTT	TTCATTAAGA	GACAACTCAC	AATTATGTAA	AAAGTGTGGT	300
T F F	S L R	D N S Q	L C K	K C G	
TTATGCCCTA	CAGGAAGCCC	TCAGAGTCCA	CCTCCCTACC	CCAGGGTCCC	350
L C P T	G S P	Q S P	P P Y P	S V P	
CTCCCCGACT	CCTTCCTCAA	CTAATAAGGA	CCCCCCTTTA	ACCCAAACGG	400
S P T	P S S T	N K D	P P L	T Q T V	
TCCAAAAGGA	GATAGACAAA	GGGGTAAACA	ATGAACCAAA	GAGTGGCAAT	450
Q K E	I D K	G V N N	E P K	S A N	
ATTCCCCGAT	TATGCCCCCT	CCAAGCAGTG	AGAGGAGGAG	AATTGGGCCC	500
I P R L	C P L	Q A V	R G G E	F G P	
AGCCAGAGTG	CCTGTACCTT	TTTCTCTCTC	AGACTTAAAG	CAAAATTAAA	550
A R V	P V P F	S L S	D L K	Q I K I	
TAGACCTAGG	TAAATTCTCA	GATAACCTTG	ACGGCTATAT	TGATGTTTTA	600
D L G	K F S	D N P D	G Y I	D V L	
CAAGGGTTAG	GACAATCCTT	TGATCTGACA	TGGAGAGATA	TAATGTTACT	650
Q G L G	Q S F	D L T	W R D I	M L L	
ACTAAATCAG	ACACTAATCC	CAAATGAGAG	AAGTGGCGCT	GTAACGTGAG	700
L N Q	T L T P	N E R	S A A	V T A A	
COCGAGAGTT	TGGCGATCCT	TGGTATCTCA	GTCAGGCCAA	CAATAGGATG	750
R E F	G D L	W Y L S	Q A N	N R M	
ACAACAGAGG	AAAGAACAAC	TCCACAGGC	CAGCAGGCAG	TTCOCAGTGT	800
T T E E	R T T	P T G	Q Q A V	P S V	

20 / 32

FIG 8 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGACCCCTCAT	TGGGACACAG	AATCAGAACA	TGGAGATTGG	TGCCACAAAC	850
D P H	W D T E	S E H	G D W	C H K H	
ATTTGCTAAC	TTGGGTGCTA	GAAGGACTGA	GCAAAAGTAG	GAAGAAGCCT	900
L L T	C V L	E G L R	K T R	K K P	
ATGAATTACT	CAATGATGTC	CACTATAACA	CAGGCAAAGG	AAGAAAATCT	950
M N Y S	M M S	T I T	Q G K E	E N L	
TACTGCTTTT	CTGGACAGAC	TAAGGGAGGC	ATTGAGGAAG	CATACCTCCC	1000
T A F	L D R L	R E A	L R K	H T S L	
TGTCACCTGA	CTCTATTGAA	GGCCAACTAA	TCTTAAAGGA	TAAGTTTATC	1050
S P D	S I E	G Q L I	L K D	K F I	
ACTCAGTCAG	CTGCAGACAT	TAGAAAAAAC	TTCAAAAGTC	TGCCTAAGCT	1100
T Q S A	A D I	R K N	F K S L	P K L	
TGGGGGCGCA	CTGGAGCACC	ACCACCACCA	CCACTGAGAT	CCGGCTGCTA	1150
A A A	L E H H	H H H	H . D	P A A N	
ACAAAGCCCC	AAAGGAAGCT	GAGTTGGCTG	GTGGCA		1186
K A R	K E A	E L A G	G		

21 / 32

FIG 9

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTTGCC	TCTCCCAATT	50
C P L C	S . S	S T G	A H C L	S Q L	
V R C	A P D P	A Q A	P I A	S P N W	
S A V	L L I	Q H R R	P L P	L P I	
GGGCTAAAGG	CTTGCCATTG	TTCCIGCACA	GCTAAGTGGC	TGGGTTTCATC	100
G . R	L A I V	P A Q	L S A	W V H P	
A K G	L P L	F L H S	. V P	G F I	
G L K A	C H C	S C T	A K C L	G S S	
CTAATCGAGC	TGAACACTAG	TCACTGGGTT	CCACGGTTCT	CTTCCATGAC	150
N R A	E H .	S L G S	T V L	F H D	
L I E L	N T S	H W V	P R F S	S M T	
. S S	. T L V	T G F	H G S	L P . P	
CCATGGCTTC	TAATAGAGCT	ATAACACTCA	CTGCATGGTC	CAAGATTCCA	200
P W L L	I E L .	H S	L H G P	R F H	
H G F	. . S Y	N T H	C M V	Q D S I	
M A S	N R A	I T L T	A W S	K I P	
TTCTTGGAA	TCCTGAGAC	CAAGAACCCC	AGGTCAGAGA	ACACAAGGCT	250
S L E	S V R P	R T P	G Q R	T Q G L	
P W N	P . D	Q E P Q	V R E	H K A	
F L G I	R E T	K N P	R S E N	T R L	
TGCCAOCATG	TTGGAAGCAG	CCACCCACCA	TTTIGGAAGC	AGCCCGCCAC	300
P P C	W K Q	P T T I	L E A	A R H	
C H H V	G S S	P P P	F W K Q	P A T	
A T M	L E A A	H H H	F G S	S P P L	
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	AGGTAACAAT	TTGGTGACCA	350
Y L G S	S G S	K D P	R . Q F	G D H	
I L G	A L G A	R T P	G N N	L V T T	
S W E	L W E	Q G P Q	V T I	W . P	
CGAAGGGACC	TGAATCCGCA	ACCATGAAGG	GATCTCCAAA	GCAATTGGAA	400
E G T	. I R N	H E G	I S K	A I G N	
K G P	E S A	T M K G	S P K	Q L E	
R R D L	N P Q	P . R	D L Q S	N W K	

22 / 32
FIG 9 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGTTCTCTCC	CAAGGCAAAA	ATGCCCCCTAA	GATGTATTCT	GGAGAATTGG	450
V P P	K A K	M P L R	C I L	E N W	
M F L P	R Q K	C P .	D V F W	R I G	
C S S	Q G K N	A P K	M Y S	G E L G	
GACCAATTIG	ACCTCAGAC	AGTAAGAAAA	AAATGACTTA	TATTCTCTG	500
D Q F D	P Q T	V R K	K .	L I F F C	
T N L	T L R Q	. E K	N D L	Y S S A	
P I .	P S D	S K K K	M T Y	I L L	
CAGTACCGCC	CTGGCCACGA	TATCTCTTTC	AAGGGGGAGA	AACCTGGCCT	550
S T A	L A T I	S S S	R G R	N L A S	
V P P	W P R	Y P L Q	G G E	T W P	
Q Y R P	G H D	I L F	K G E K	P G L	
CCTGAGGGAA	GTATAAATTA	TAACACCATC	TTACAGCTAG	ACCTGTTTIG	600
. G K	Y K L	. H H L	T A R	P V L	
P E G S	I N Y	N T I	L Q L D	L F C	
L R E V	. I I	T P S	Y S .	T C F V	
TAGAAAAGGA	GGCAAATGGA	GTGAAGTGCC	ATATTIACAA	ACTTCTTTT	650
. K R R	Q M E	. S A	I F T N	F L F	
R K G	G K W S	E V P	Y L Q	T F F S	
E K E	A N G	V K C H	I Y K	L S F	
CATTAAAGA	CAACTGCAA	TTATGTTAAC	AGTGTGATTT	GTGTCTCTAC	700
I K R	Q L A I	M L T	V .	F V F L H	
L K D	N S Q	L C .	Q C D L	C S Y	
H .	K T	T R N	Y V N	S V I C V P T	
ACGGAAGCCC	TCAGATTCTA	CTCCCCACCC	CCGGCATCTC	CCCTGAATCC	750
G S P	Q I L	L P T P	G I S	P E S	
T E A L	R F Y	S P P	P A S P	L N P	
R K P	S D S T	P H P	R H L	P . I P	
CTCCCCAACT	TATT				764
L P N L					
S P T Y					
P Q L I					

23 / 32

FIG 10

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCCTGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATIGGC	TCTCCCAATT	50
C P L C S . S	S T G A H C L	S Q L			
V R C A P D P	A Q A P I A	S P N W			
S A V L L I	Q H R R	P L P L P I			
GGGCTAAAGG	CTTGCCATTG	TTCTTGCACA	GCTAAGTGGC	TGGGTTCATC	100
G . R L A I V	P A Q L S A	W V H P			
A K G L P L	F L H S . V	P G F I			
G L K A C H C	S C T A K C L	G S S			
CTAATCGAGC	TGAACACTAG	TCACITGGGTT	CCACGGTCTCT	CTTCCATGAC	150
N R A E H .	S L G S T V L	F H D			
L I E L N T S	H W V P R F S	S M T			
. S S . T L V	T G F H G S	L P . P			
CCATGGCTTC	TAATAGAGCT	ATAACACTCA	CTGCATGGTC	CAAGATTCCA	200
P W L L I E L .	H S L H G P	R F H			
H G F . . S Y	N T H C M V	Q D S I			
M A S N R A	I T L T A W S	K I P			
TTCTTIGGAA	TCCGTGAGAC	CAAGAACCCC	AGGTACAGAGA	ACACAAGGCT	250
S L E S V R P	R T P G Q R	T Q G L			
P W N P . D	Q E P Q V R E	H K A			
F L G I R E T	K N P R S E N	T R L			
TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTIGGAAGC	GGCCCCGCCAC	300
P P C W K Q	P T T I L E A	A R H			
C H H V G S S	P P P F W K R	P A T			
A T M L E A A	H H H F G S	G P P L			
TATCTTIGGA	GCTCTGGGAG	CAAGGAACCCC	CAGGTACACAA	TTTGGTGAAC	350
Y L G S S G S	K D P Q V T I	W . P			
I L G A L G A	R T P R . Q	F G D H			
S W E L W E	Q G P P G N N	L V T			
ACGAAGGGAC	CTGAATCCGC	AACCATGAAG	GGATCTCCAA	AGCAATTGGA	400
R R D L N P Q	P . R D L Q	S N W K			
E G T . I R	N H E G I S K	A I G			
T K G P E S A	T M K G S P K	Q L E			

24 / 32
FIG 10 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AATGTTCCCTC	CCAAGGCAAA	AATGCCCCCTA	AGATGTATTTC	TGGAGAATTG	450
C S S	Q G K	N A P K	M Y S	G E L	
N V P P	K A K	M P L	R C I L	E N W	
M F L	P R Q K	C P .	D V F	W R I G	
GGACCAATCT	GACCCCTCAGA	CAGTAAGAAA	AAAAATGACT	TATATTCTTC	500
G P I .	P S D	S K K	K N D L	Y S S	
D Q S	D P Q T	V R K	K M T	Y I L L	
T N L	T L R	Q .	E K K .	L I F F	
TGCAGTACCG	CCTGGCCACG	GATATCCTCT	TCAAGGGGGA	GAAACCTGGC	550
A V P	P G H G	Y P L	Q G G	E T W P	
Q Y R	L A T	D I L F	K G E	K P G	
C S T A	W P R	I S S	S R G R	N L A	
CTCCTGAGGG	AAGTATAAAT	TATAACACCA	TCTTACAGCT	AGACCTGTTT	600
P E G	S I N	Y N T I	L Q L	D L F	
L L R E	V .	I I T P	S Y S .	T C F	
S .	G K Y K L	. H H	L T A	R P V L	
TGTAGAAAAG	GAGGCAAATG	GAGTGAAGTG	CCATATTTAC	AAACTTTCTT	650
C R K G	G K W	S E V	P Y L Q	T F F	
V E K	E A N G	V K C	H I Y	K L S F	
. K R	R Q M E .	S A	I F T	N F L	
TTCATTAAAA	GACAACTCGC	AATTATGTAA	ACAGTGTGAT	TGTGTGCCIA	700
S L K	D N S Q	L C K	Q C D	L C P T	
H .	K T T R	N Y V N	S V I	C V L	
F I K R	Q L A	I M .	T V .	F V S Y	
CAGGAAGCCC	TCAGATCTAC	CTCCCTACCC	CGGCATCTCC	CTGACTCCTT	750
G S P	Q I Y	L P T P	A S P .	L L	
Q E A L	R S T	S L P	R H L P	D S F	
R K P	S D L P	P Y P	G I S	L T P S	
CCCCAACTAA	TAAGGACCCA	CTTCAGCCCCA	AACAGTCCAA	AAGGACATAG	800
P Q L I	R T H	F S P	N S P K	G H	
P N .	. G P T	S A Q	T V Q	K D I	
P T N	K D P	L Q P K	Q S K	R T .	

25 / 32
FIG 11

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GGCATTGATA	GCACCCATCA	GATGGCCAAA	TCATTATTTA	CTGGACCAGG	50
G I D S	T H Q	M A K	S L F T	G P G	
A L I	A P I R	W P N	H Y L	L D Q A	
H . .	H P S	D G Q I	I I Y	W T R	
CCTTTTCAAA	ACTATCAAGC	AGATAGGGCC	CGTGAAGCAT	GCCAAAGAAA	100
L F K	T I K Q	I G P	V K H	A K E I	
F S K	L S S R	. G P	. S M	P K K	
P F Q N	Y Q A	D R A	R E A C	Q R N	
TAATCCCTG	CCTTATCGCC	ATGTTCTTTC	AGGAGAACAA	AGAACAGGCC	150
I P C	L I A	M F L Q	E N K	E Q A	
. S P A	L S P	C S F	R R T K	N R P	
N P L	P Y R H	V P S	G E Q	R T G H	
ATTACCCAGG	GGAAGACTGG	CAACTAGATT	TTACCCACAT	GGCCAAATGT	200
I T Q G	K T G N	. I	L P T W	P N V	
L P R	G R L A	T R F	Y P H	G Q M S	
Y P G	E D W	Q L D F	T H M	A K C	
CAGGGATTTC	AGCATCTACT	AGTCIGGGCA	GATACTTTCA	CTGGTIGGGT	250
R D F	S I Y .	S G Q	I L S	L V G W	
G I S	A S T	S L G R	Y F H	W L G	
Q G F Q	H L L	V W A	D T F T	G W V	
GGAGTCTTCT	CCTTGTAGGA	CAGAAAAGAC	CCAAGAGGTA	ATAAAGGCAC	300
S L L	L V G	Q K R P	K R .	. R H	
G V F S	L . D	R K D	P R G N	K G T	
E S S	P C R T	E K T	Q E V	I K A L	
TAATGAAATA	ATTCCAGAT	TTCGACTTCC	CCCAGGATTA	CAGGGTGACA	350
. . N N	S Q I	W T S	P R I T	G . Q	
N E I	I P R F	G L P	P G L	Q G D N	
M K .	F P D	L D F P	Q D Y	R V T	

26 / 32
FIG 11 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCCCCGC	TTTCAAGGCT	GCAGTAAACC	AGGGAGTATC	CCAGGTGTTA	400
W P R	F Q G C	S N P	G S I	P G V R	
G P A	F K A	A V T Q	G V S	Q V L	
M A P L	S R L	Q . P	R E Y P	R C .	
GGCATACAAT	ATCACTTACA	CTGTGCCCTGG	AGGCCACAAT	CCTCCAGAAA	450
H T I	S L T	L C L E	A T I	L Q K	
G I Q Y	H L H	C A W	R P Q S	S R K	
A Y N	I T Y T	V P G	G H N	P P E K	
AGTCAAGAAA	ATGAATGAAA	CACTCAAAGA	TCTAAAAAAG	CTAACCCAAG	500
S Q E N	E . N	T Q R	S K K A	N P R	
V K K	M N E T	L K D	L K K	L T Q E	
S R K	. M K	H S K I	. K S	. P K	
AAACCCACAT	TGCATGACCT	GTTCGTGTGC	CTATAACCTT	ACTAAGAATC	550
N P H	C M T C	S V A	Y N L	T K N P	
T H I	A . P	V L L P	I T L	L R I	
K P T L	H D L	F C C	L . P Y	. E S	
CATAACTATC	CCCCAAAAG	CAGGACTTAG	CCCATACGAG	ATGCTATATG	600
. L S	P K K	Q D L A	H T R	C Y M	
H N Y P	P K S	R T .	P I R D	A I W	
I T I	P Q K A	G L S	P Y E	M L Y G	
GATGGCCTTT	CCTAACCAAT	GACCTTGTC	TTGACTGAGA	AATGGCCAAC	650
D G L S	. P M	T L C	L T E K	W P T	
M A F	P N Q .	P C A	. L R	N G Q L	
W P F	L T N	D L V L	D . E	M A N	
TTAGTTCAG	ACATCACCTC	CTTAGCCAAA	TATCAACAAG	TTCTTAAAAC	700
. L Q	T S P P	. P N	I N K	F L K H	
S C R	H H L	L S Q I	S T S	S . N	
L V A D	I T S	L A K	Y Q Q V	L K T	

27 / 32
FIG 11 (continued)

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
<hr/>				
ATCACAGGGA ACCTGTCCCC GAGAGGAGGG AAAGGAACTA TTCCACCCCTG				
750				
H R E P V P E R R E R N Y S T L				
I T G N L S P R G G K G T I P P W				
S Q G T C P R E E G K E L F H P G				

GTGACATG

758

V T

. H

D M

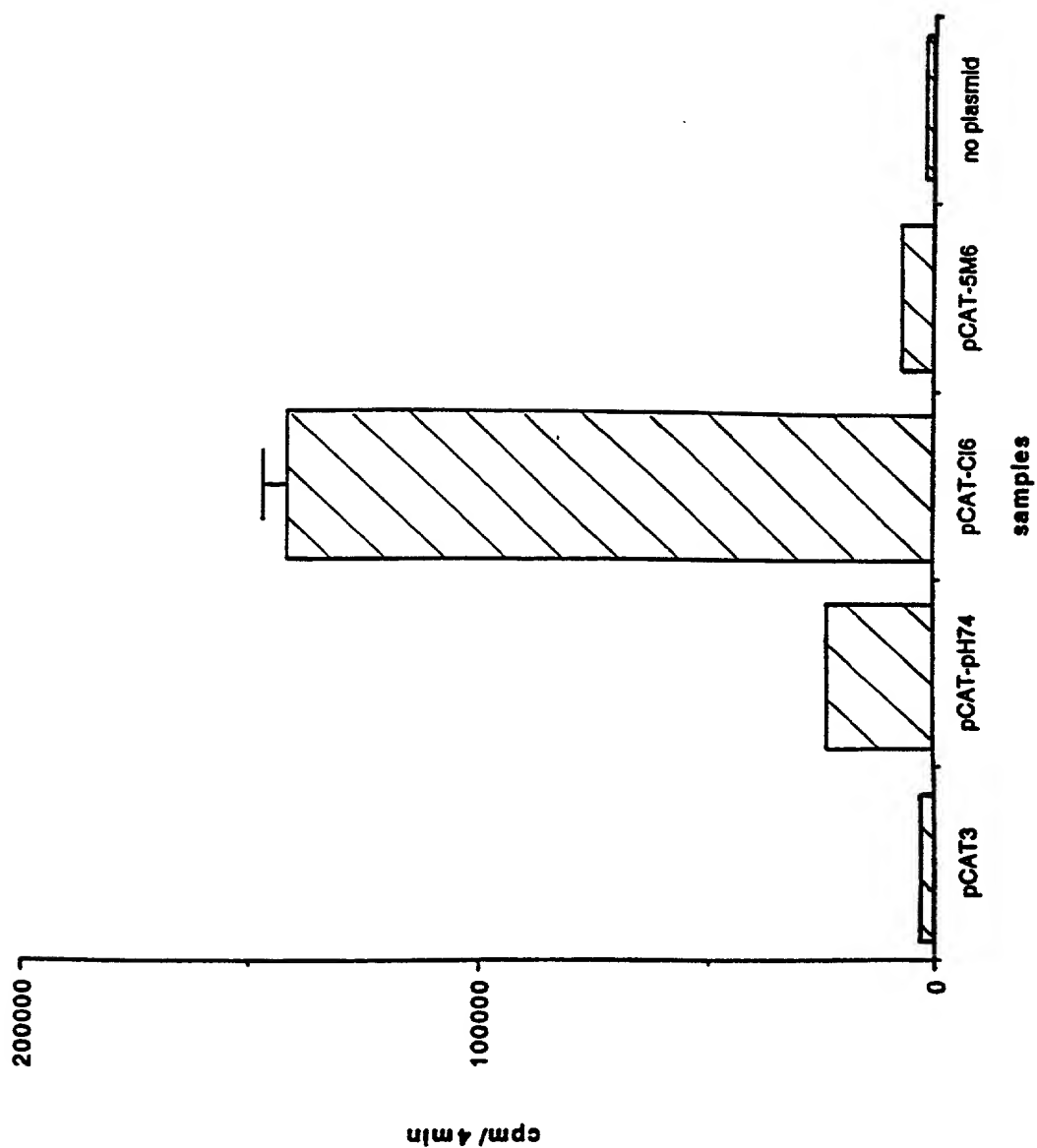


FIG 12

100	ATGGCCCTC CTATACGAC TTCTCTCTT ACCTCTCTT TACCCCTTT GCGCCCTC CATCTCTG TCACACGAT AGCTCCCTT	34
200	HA L P Y H T F L P T V L L P P F A L T A P P P C C C T T T B S S P Y	67
300	ACCAGAGT TCATGACG ACCGCGCTC CTGGAAAT TCATGACCA TCATATGGA GTTATCTTA GCGAACTCC ACCTCTCTC CCGACACCA	100
400	Q E F L . R T R L P G N I D A P S Y R S L S K G N S T F T A H T H	134
500	TATCCCGC AACTCTATA ACTCTCCAC TCTTGCA TCATATGGA GTTATCTTA GCGAACTCC ACCTCTCTC CCGACACCA	167
600	M P R N C Y N S A T L C H A N T H Y W T O K H I P P S C P O G L	200
700	GAGCCACT TCCTCTCAC CATACGATA TGTCTGAG GCGCGAAT CAAGCTCAG CAGAGAAA ACAAGTAA GAGCAATCT	234
800	G A T V C M T Y F T H T S M S D G G G I Q G O A R E K Q V K E A I S	267
900	CCGACCTAC CCGGACAT AGCAGCTCA GCGCTTCA AGAATGAT CTCTCAAA TCATGAA TCATGAA CTCTCTCAC CATCTCCG TGTGAGCT	300
1000	Q L T R G H S T P S P Y K G L V L S K L H E T L R T H T R L V S L	334
1100	ATTAATAC ACCTCTCTC GCGTCCAG GGTCTCAG CAAACCTA CTATCTGAT GATCTCTC CCGCTCTC TCAGCCATA CATTCATC	367
1200	F N T T L T R L H E V S A Q N P T N C W H C L P L H F R P Y I S I	400
1300	CTCTCTCTC AACTCTCAC CATCTCTAC ACAGAAAT ACAGCTCTC GTTATGTA GAGCTCTT TTCTCTCTC CCGAACTAC CATCTCTCA	434
1400	P V P E Q W N M F S T E I N T S V L V G P L V S N L E I T H T S	467
1500	ACCTCTCTC TGTAAATTT AGCAATCTA TAGACACAC CAGCTCCCA TGCATGAT GGTCTCTC CCGCTCTC TCAGCCATA CATTCATC	500
1600	L C V K P S N T I D T T S S Q C I R W V T P P T R I V C L P S G	534
1700	ACTCTCTC GTCTCTCTA CTCTCTCTA TCATCTCTC AGAATCTT GAGATCTT GAGATCTT TGTCTCTC TGTCTCTC GAGATCTT	567
1800	I F P V C G T S A Y H C L N G S S E S H C F L S F L V P P M T I Y	600
1900	ACTGACAG ATTATACAA TCATCTCTA CTTAGCTCC ACAGAAAT AGTACCTT CTCTCTCTC TTCTCTCTC AGAATCTA CCGACCTAC	634
2000	T E Q D L Y N H V V P K P H N K R V P I L P F V I R A G V L G R L G	667
2100	GTCTCTCTC TGTCTCTA ACAGCTCTA CTCTCTCTA CTCTCTCTA TCATCTCTC CCGAACTAC CATCTCTCA	700
2200	T G I S I T T S T Q F Y Y K L S Q E I N G D H E Q V T D S L V T	734
2300	CTCTCTCTC CCGACCTAC AGTCTCTT CAACTCTA GAGCTCTA CTCTCTCTC GCGAACTAC CCGAACTAC TTCTCTCTC	767
2400	L Q D Q L N S L A A V V L Q N R R A L D L L T A K R G G T C L F L	800
2500	GAGAGAC GCTCTCTA TGTCTCTA TCATCTCTA AGTCTCTA AGTCTCTA TCATCTCTC CCGAACTAC CATCTCTCA	834
2600	G E R C Y Y V N U S R I V T E K V K E I R D R I Q C R A E E L Q N	867
2700	ACAGACAG CTCTCTCTC CTCTCTCTC GGTCTCTC TTCTCTCTC CTCTCTCTC TTCTCTCTC TTCTCTCTC TTCTCTCTC	900
2800	T E R M G L L S Q W M P W V L P F L G P L A A L I L L L L L F G P C	934
2900	TATCTCTA CTCTCTCTC TTCTCTCTC GAGCTCTA AGTCTCTA GGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA	967
3000	I F N L L V K F V S S R I E A V K L Q H V L Q H E P Q H E S H T K	1000
3100	ATCTCTCTC GAGCTCTC CCGCTCTC AGCTCTCTC CCGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC	1034
3200	I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L	1067
3300	TATCTCTA TCATCTCTA AGCTCTCTA GCGCTCTA CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC	1100
3400	C P N S A G S S	1134
3500	TATCTCTA TCATCTCTA AGCTCTCTA GGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA	1167
3600	I F N L L V K F V S S R I E A V K L Q H V L Q H E P Q H E S H T K	1200
3700	ATCTCTCTC GAGCTCTC CCGCTCTC AGCTCTCTC CCGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC	1234
3800	I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L	1267
3900	TATCTCTA TCATCTCTA AGCTCTCTA GCGCTCTA CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC	1300
4000	C P N S A G S S	1334
4100	TATCTCTA TCATCTCTA AGCTCTCTA GGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA	1367
4200	I F N L L V K F V S S R I E A V K L Q H V L Q H E P Q H E S H T K	1400
4300	ATCTCTCTC GAGCTCTC CCGCTCTC AGCTCTCTC CCGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC	1434
4400	I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L	1467
4500	TATCTCTA TCATCTCTA AGCTCTCTA GCGCTCTA CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC	1500
4600	C P N S A G S S	1534
4700	TATCTCTA TCATCTCTA AGCTCTCTA GGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA	1567
4800	I F N L L V K F V S S R I E A V K L Q H V L Q H E P Q H E S H T K	1600
4900	ATCTCTCTC GAGCTCTC CCGCTCTC AGCTCTCTC CCGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC	1634
5000	I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L	1667
5100	TATCTCTA TCATCTCTA AGCTCTCTA GCGCTCTA CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC	1700
5200	C P N S A G S S	1734
5300	TATCTCTA TCATCTCTA AGCTCTCTA GGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA	1767
5400	I F N L L V K F V S S R I E A V K L Q H V L Q H E P Q H E S H T K	1800
5500	ATCTCTCTC GAGCTCTC CCGCTCTC AGCTCTCTC CCGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC	1834
5600	I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L	1867
5700	TATCTCTA TCATCTCTA AGCTCTCTA GCGCTCTA CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC	1900
5800	C P N S A G S S	1934
5900	TATCTCTA TCATCTCTA AGCTCTCTA GGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA	1967
6000	I F N L L V K F V S S R I E A V K L Q H V L Q H E P Q H E S H T K	2000
6100	ATCTCTCTC GAGCTCTC CCGCTCTC AGCTCTCTC CCGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC	2034
6200	I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L	2067
6300	TATCTCTA TCATCTCTA AGCTCTCTA GCGCTCTA CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC	2100
6400	C P N S A G S S	2134
6500	TATCTCTA TCATCTCTA AGCTCTCTA GGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA	2167
6600	I F N L L V K F V S S R I E A V K L Q H V L Q H E P Q H E S H T K	2200
6700	ATCTCTCTC GAGCTCTC CCGCTCTC AGCTCTCTC CCGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC GAGCTCTC	2234
6800	I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L	2267
6900	TATCTCTA TCATCTCTA AGCTCTCTA GCGCTCTA CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC CCGCTCTC	2300
7000	C P N S A G S S	2334

Poly A signal

Cap site

AATCTCTA TCATCTCTA AGCTCTCTA GGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA AGTCTCTA

FIG13

FIG 14

100 CAGCAACCCC CTTTGGGTCC CCTCCCATG TATGGAGCT CTGTTTTCAC TCATTTCAC TCATTAATTC CATGCAACTG CACTCTCTCG GTCCGTGTTT
 200 TTATGGCTC AAGCTGAGCT TTGTTGGCC ATCCACCACT GCTGTTGGCC ACCGCTGCT GACTTCCATC CTTTGGATC CAGCAGATG
 300 TCCGCTGTG TCTGATCCA GCACAGGCGC CCATGGCTC TCCCAATGG GCTAAAGGCT TGCCATGTTT CCTGCACAGC TAAGTGCCCTG GGTTCATCCT
 400 AAJGAGCTG AACACTAGTC ACTGGGTCC ACGTCTCTT TCCATGACCC ATGGCTTCTA ATAGAGCTAT AACACTACT GCATGGTCCA AGATTCATTT
 500 CCTTGAATC CBTGAGACCA AGAACCCAG GTCAGAGAAC ACAAGGTG CACCATGTT GGAAGCAGCC CACCACTTT TGAAGAGCAG CCGGCCACTA
 600 TCTTGGAGC TCTGGGAGCA AGGACCCAG GTAACATTT GGTGACCAAG AGGGACCTG AATCCGCAAC CATGAAGGGA TCTCCAAGC
 3 M G N
 700 GTTCCCCCG AGCAAAAT GCCCCTAGAA CGTATTCTGG AGATTTGGG CCAATGTGAC ACTCAGAGC TAAGAAAGAA AGATTTATA TTCTCTGCA
 37 V P P E A K M P L E R I L E N W D Q C D T Q T L R K K R F I F C S
 800 GTACGCTG GCCACAATAT CCTCTTCAAG GGAGAGAAC CTGGCTTCT GAGGGAAGTA TAAATATATA CATCATCTTA CAGTAGACC TCCTCTGTAG
 90 T A W P Q Y P L Q G R E T W L P E G S I N Y N I I L Q L D L F C R
 1000 AAAGGAGGC AATGGAGTG AAGTGCCATA TGTGCAAACT TCTTTTCAT TAAGAGACAA CTCACAATTA TGTAAAGT GTGGTTTATG CCTACAGGA
 103 K E G K W S E V P Y V Q T F F S L R D N S Q L C K K C G L C P T G
 1000 AGCCTCAGA GTCCACCTCC CTACCCCTCC GTCCCTCTTC CGACTCTTC CTCACCTAAT AAGGACCCCT CTTTAAACCA AACGGTCCA AAGGAGATAG
 137 S P Q S P P Y P S V P S P T P S S T N K D P P L T Q T V Q T V Q K E I D
 1100 ACNAGGGT AAACAATGAA CCAAGAGTG CCAATATTC CCGATTATG CCGTCCCAAG CAGTGAGAGG AGGAGATTC GSCCAGCCA GAGTGCTGT
 170 K G V N N E P K S A N I P R L C P L Q A V R G G E F G P A R V P V
 1200 ACCTTTTCT CTCCTCAGCT TAAAGCAAT TAAATAGAC CTAGTAAAT TCTCAGATA CCTGAGGC TATATTGATG TTTTACAGG GTTAGACAA
 203 P F S L S D L K Q I K I D L G K F S D N P D G Y I D V L Q G L G Q
 1300 TCCTTGATC TGACATGGAG AGATATAATG TTTACTATA ATCAGACACT AACCCCAAT GAGAGAAGTG CCGCTGTAC TGCAGCCCGA GAGTTGGCG
 237 S F D L T W R D I M L L N Q T L T P N E R S A A V T A A R E F G D
 1400 ATCTTGGTA TCTCAGTCAG GCCAACATA GGATGACAA AGAGGAAGA ACACTCCA CAGGCCAGCA GGCAGTTCCC AGTGTAGACC CTCATTGGGA
 270 L W Y L S Q A N N R M T T E E R T T P T G Q Q A V P S V D P H W D
 1500 CACAGATCA GAACATGGAG ATTGGTCCA CAACATTTG CTACTTGG TGTAGAAG ACTGAGGAA ACTAGGAAGA AGCTATGAA TTACTCAATG
 303 T E S E H G D W C H K H L L T C V L E G L R K T R K K P M N Y S M
 1600 ATGTCCACTA TAACACAGG AAAGGAAGAA AATCTTACTG CTTTCTGGA CAGACTAAG GAGGCATTA GGAAGCATAC CTCCTGTCA CCTGACTTA
 337 M S T I T Q G K E E N L T A F L D B L R E A L R K H T S L S P D S I
 1700 TTGAAGGCA ACTAATCTTA AAGGATAAGT TTATCACTA GTCAGTGA GATCTAGAA AAAACTTCA AAGTCCGTC TTAGGCTCG AACAAACTT
 370 E G Q L I L K D K F I T Q S A A D I R K K L Q K S V L G S E Q N L
 1800 AGAAACCTTA TTGAACCTGG CAACCTCGT TTTTATAT AGAGTACAG AGGACAGC AGATGGGATA AAAAAGG GGCACCGCT
 403 E T L L N L A T S V F Y N R D Q E E Q A E W D K W D K K K R A T A
 1900 TTATGATG CCTCAGGCA AGGGACTTT GGAGGCTCTG GAAAGGGA AAGCTGGCA AATAGGAGC CTAATAGGC TTGCTTCCAG TCGGTCTAC
 437 L V M A L R Q A D F G G S G K G K S W / A N R K P N R A C F Q C G L Q
 2000 AAGGACCTT TAAAGAAT TGTCCAATA GAATAGCC GCCCTTGT CCATGCCCT TAGCTCAGG GAATCACTG AAGGCCACT GCCCAGGG
 470 G H F K K D C P N R N K P P C / R P C P L R Q G N / H W K A H C P R G
 2055 ATCAAGATAC TCTGATCAG AACCATTA CCAGATGATC CAGACGAG ACTGA
 487 S R Y S E S E A I N Q M I Q Q Q D

FIG 15

100 GGACCCGTAG TATGGGGTAA TCCCTCCCG GAAACCAAGC CCCAGTACTC AGAAGAAGAA ATAGAATGGG GAACTCAGC AGGACATGGT TTCTCCCT
 34 G | P V V W G N P L R E T K P Q Y S E E E I E W G T S R G H G P L P S
 200 CAGGATGGCT AGCCACTGAA GAAGGAAAAA TACTTTTGCT GGCAGCTAAC CAATGGAAAT TACTTAAAC CCTTCAGCAA ACCTTCCACT TAGGCATTGA
 67 G W L A T E E G K I L L L A A N Q W K L L K T L Q Q T P E L G I D
 300 TAGCACCCAT CAGATAGCCA AATCATTTAT TACTGGACCA GGCCTTTTCA AACTATCAA GCAGATAGTC AGGCGCTGTG AAGTGTGCGA AAGAAATAT
 100 S T E Q I A K S L P T G P G L P K T I K Q I V R A E E V E Q R N N
 400 CCCCCTGCTT ATCGCCAAAGC TCCTTCAGGA GAACAAAGAA CAGGCAATTA CCCAAGAGAA GACTGGCAAC TAGATTTTAT CCACATGCGA AAATCACAGG
 134 P L P Y R Q A P S G E Q R T G N Y P R E D W Q L D F I H M P K S Q G
 500 GATTCAGTG TCTACTAGTC TGGGTAGATA CTTTCACTGG TTGGGCAGAG GCCTTCCCT GTAGGACAGA AAGTTCCAA GAGGTATATA AGGCACTAGT
 167 F Q C L L Y W V D T F T G W A E A P P C R T E K P Q E V I K A L V
 600 TCATGAAGTA ATTCCAGAT TCGACTTCC CTGAGGCTTA CAGAGTGACA ATGTCTCTGC TTTCAGGCC ACAGTAACCC AGGAGTATC CCAGGCGTTA
 200 H E V I P R F G L P . G L Q S D N G P A F K A T V T Q G V S Q A L
 700 GGTATAGAAAT ATCACTTACA CTGCACCTAG AGGCCACAAT CCTCAGGGAA CGTTGAGAA ATGA~~AT~~ACAC TCAACGACA TCTAAGCAG CTAAACCCAGG
 234 Q I E Y H L H C T . R P Q S S G K V E K M K T L K R H L N K L T Q E
 800 AAACCCAGCT CGCATGGTCT GCTCTGTGT CTATAGCTT ACTAAGATC CAAACTCTC CCCAAAGGC AGGACTTAGC CCATACAGAA TGCTGTATGG
 267 T H L A W S A L L S I A L L R I Q N S P Q K A G L S P Y R M L Y G
 900 ACGGTCTTC CTAAACCAATG ACCTTCTCT TGAACCAAG ATGGCCAAT TGTGTGAGA CATCACTCC TTAGCCAAAT ATCAACAAAT TCTTAAACA
 300 R S F L T N D L L L D Q E M A N L V A D I T S L A K Y Q Q V L K T
 1000 TTACAGAGG CCTGTCCCG AGAGGAGGA AAGAAATAT TCCACCTGG TGTATAGTA TTATCAAGT CCCTTCCCTC TAATTCCTCA TCCTTAGACA
 134 L Q G A C P R E E G K E I P H P G V M V L V K S L P S M S P S L D T
 1100 CATCTGGG AGGACCTTAC CCAGTCATTT TATCTATCC AACTGCGTT AAGTGGCTG GAGTGGAGTC TTGATACAT CACACTGGA TCAACCTG
 367 S W G G P Y P V I L S I P T A V K V A G V E S W I H H T R I K P W
 1197 GATCTGCCG AAGGAACCCG AAAATCCAGG GGACAAAGCT AGCTATTTCT TTGAACCTCT AGAGATCTG TGCTGTCT TCAAGCAACA ACCGTGA
 398 I L P K E P E N P G D N A S Y F F E P L E D L C L L F K Q Q P .

100 GAGAGGCA GATATGATG GATGAGCA AGAGAGCA GTATAGAA AAGAGAA GTCAGCAA GAATAGCA GAGAGAA
 E N S S I S W L A E V G K D S K K . R K K G E S Q R K K R E E E T
 200 OAGAGGAA GTAGAGGA GAAGAGTA GTATAGAA AAGAGGAC GATATGATG TATAGGATG GTATATGTC TGTATGTA GCGAGGAT
 K K N L K R E R S S K E K T V Y P I P L K A R V N F C L P S Q G I
 300 ATCTGCTA TGTAGGAT GAGAGGAT GTCGTCGTC ACTATGCA GAGAGGAG AGCTGATG TTTCTAGTC GCGAGGATTA GATGCGCA
 F F L C G T S T Y I C L P T N W T G T R T L V F L S P N I N I A P
 400 GAGAGGCA GATATGATG GATAGGAT GTCGTCGTC ACTATGCA GAGAGGAG AGCTGATG TTTCTAGTC GCGAGGATTA GATGCGCA
 G N Q T L L V P V K A K V R Q C R A I Q L I S L F I G L G M A T A T
 500 GAGAGGAT AAGAGGAT TTTCTAGTC GATATGCA GTATAGCA AATGAGCA GATGAGCA GATGAGCA AATGAGCA
 G T G I A G L S T S L S Y Y H T L S K N F S D S L Q E I M K S I L
 600 TATGAGCA TGTAGGAT GATGAGGAT GCGAGGAT GCGAGGAT GCGAGGAT GCGAGGAT GCGAGGAT GCGAGGAT
 T L Q S Q L D S L A A M T L Q N R R G P H L L T A F K G G L C T F
 700 TATGAGGAT AAGAGGAT TTTCTAGTC GATAGGAT TGTAGGAT GATGAGGAT AAGAGGAT TGTAGGAT GATGAGGAT
 L G E C C F Y T N Q S G I V R D A T W H L Q E R A S D I R Q C L S
 800 GAGAGGAT TGTAGGAT GATGAGGAT TTTCTAGTC GATAGGAT TGTAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 N S Y T N L W S W A T W L L P F L G P M A A I L L L L T F G P C I
 900 TTTCTAGTC GATGAGGAT TGTAGGAT TGTAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 F K L L V K F V S S R I E A I K L Q M V L Q M E P Q M S S T N N F
 1000 TATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 Y Q G P L E R S T G T S T S L E I P L W K T L Q L Q G P F F A P I Q
 1100 GAGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 Q E V A R A V I G Q I P N S S W G V L F R G G I E E . A C W Q P
 1200 TATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 H S P R W I S V P P Q P W C P L W P C L R S P S A C H C T V G A S
 1300 TATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 F W A G Q G R S Q L P Q L A G R Y G G R D A G G N Q G C A W R L R A
 1400 GAGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 S M S S R W A W A R R A P H S G S E G L S T W A R Q M L C S T S S
 1500 GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 L G L S C L P R G A G L R E H A A C P C L S P P P R R G F L H S P
 1600 GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 S F P D K H H P L S T V P S P I N H P R V E E C G H T A R D W Q A V
 1700 TATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT GATGAGGAT
 P L A A L V R D P L R E A S W A P E S G G D L E N L Y V L R D C
 1719 TATGAGGAT GATGAGGAT
 K Y T N Q H

FIG 16

**DECLARATION AND POWER OF ATTORNEY
UNDER 35 USC §371(c)(4) FOR
PCT APPLICATION FOR UNITED STATES PATENT**

As a below named inventor, I hereby declare that:
my residence, post office address and citizenship are as stated below under my name;

I verily believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought, namely the invention entitled: RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES described and claimed in international application number PCT/FR98/01460 filed July 7, 1998.

I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations §1.56.

Under Title 35, U.S. Code §119, the priority benefits of the following foreign application(s) filed within one year prior to my international application are hereby claimed:

French Patent Application No. 97 08816 filed July 7, 1997

The following application(s) for patent or inventor's certificate on this invention were filed in countries foreign to the United States of America either (a) more than one year prior to my international application, or (b) before the filing date of the above-named foreign priority application(s):

I hereby appoint the following as my attorneys of record with full power of substitution and revocation to prosecute this application and to transact all business in the Patent Office:

**James A. Oliff, Reg. No. 27,075; William P. Berridge, Reg. No. 30,024;
Kirk M. Hudson, Reg. No. 27,562; Thomas J. Pardini, Reg. No. 30,411;
Edward P. Walker, Reg. No. 31,450; Robert A. Miller, Reg. No. 32,771;
Mario A. Costantino, Reg. No. 33,565; and Caroline D. Dennison, Reg. No. 34,494.**

ALL CORRESPONDENCE IN CONNECTION WITH THIS APPLICATION SHOULD BE SENT TO OLIFF & BERRIDGE, PLC, P.O. BOX 19928, ALEXANDRIA, VIRGINIA 22320, TELEPHONE (703) 836-6400.

I hereby declare that I have reviewed and understand the contents of this Declaration, and that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

1 **Typewritten Full Name
of Sole or First Inventor**

Glauca PARANHOS-BACCALA
Given Name Middle Initial Family Name

2 **Inventor's Signature**

Glauca Paranhos Baccala

3 **Date of Signature**

june 28 1998
Month Day Year

Residence:

Lyon
City

State or Province

FRANCE
Country

Citizenship:

BRASIL

Post Office Address:

75 Cours Gambetta

(Insert complete mailing
address, including country)

69003 Lyon FRANCE

Note to Inventor: Please sign name on line 2 exactly as it appears in line 1 and insert the actual date of signing on line 3.

IF THERE IS MORE THAN ONE INVENTOR USE PAGE 2 AND PLACE AN "X" HERE ☒
(Discard this page in a sole inventor application)

2-00 1 **Typewritten Full Name
of Joint Inventor**

Florence KOMURIAN-PRADEL
Given Name Middle Initial Family Name

2 **Inventor's Signature:**

Florence PRADEL

3 **Date of Signature:**

21 JUIN 1999

Residence:

Poleymieux Au Mont D'Or

City

State or Province

Year
FRANCE FRX
Country

Citizenship:

FRANCE

Post Office Address:

114 Chemin du Pavillon

(Insert complete mailing
address, including country)

69250 Poleymieux Au Mont D'Or FRANCE

3-00 1 **Typewritten Full Name
of Joint Inventor**

Frederic BEDIN
Given Name Middle Initial Family Name

2 **Inventor's Signature:**

Frederic BEDIN

3 **Date of Signature:**

30 JUIN 1999

Residence:

Lyon

City

State or Province

Year
FRANCE FRX
Country

Citizenship:

FRANCE

Post Office Address:

6 Rue Gaspard Andre

(Insert complete mailing
address, including country)

69002 Lyon FRANCE

4-00 1 **Typewritten Full Name
of Joint Inventor**

Mireille SODOYER
Given Name Middle Initial Family Name

2 **Inventor's Signature:**

Mireille SODOYER

3 **Date of Signature:**

30 JUIN 1999

Residence:

Sainte Foy Les Lyon

City

State or Province

Year
FRANCE FRX
Country

Citizenship:

FRANCE

Post Office Address:

5 rue du Brulet

(Insert complete mailing
address, including country)

69110 Sainte Foy Les Lyon FRANCE

5-00 1 **Typewritten Full Name
of Joint Inventor**

Catherine OTT
Given Name Middle Initial Family Name

2 **Inventor's Signature:**

Catherine OTT

3 **Date of Signature:**

30 JUIN 1999

Residence:

Lyon

City

State or Province

Year
FRANCE FRX
Country

Citizenship:

FRANCE

Post Office Address:

103 Avenue Berthelot

(Insert complete mailing
address, including country)

69007 Lyon FRANCE

Note to Inventor: Please sign name on line 2 exactly as it appears in line 1 and insert the actual date of signing on line 3.

This form may be executed only when attached to the first page of the Declaration and Power of Attorney of the application to which it pertains.

IF THERE IS MORE THAN ONE INVENTOR USE PAGE 2 AND PLACE AN "X" HERE ☒
(Discard this page in a sole inventor application)

6-00
1 **Typewritten Full Name
of Joint Inventor**

Francois MALLET
Given Name Middle Initial Family Name

2 **Inventor's Signature:**

Francois Mallet

3 **Date of Signature:**

30 Jun 1999
Month Day Year

Residence:

Villeurbanne FRANCE FRX
City State or Province Country

Citizenship:

FRANCE

Post Office Address:

84 rue Anatole France

(Insert complete mailing
address, including country)

69100 Villeurbanne FRANCE

7-00
1 **Typewritten Full Name
of Joint Inventor**

Herve PERRON
Given Name Middle Initial Family Name

2 **Inventor's Signature:**

Herve Jean-François Perron

3 **Date of Signature:**

9 Août 1999
Month Day Year

Residence:

Lyon FRANCE FRX
City State or Province Country

Citizenship:

FRANCE

Post Office Address:

15 rue de Boyer

(Insert complete mailing
address, including country)

69005 Lyon FRANCE

8-00
1 **Typewritten Full Name
of Joint Inventor**

Bernard MANDRAND
Given Name Middle Initial Family Name

2 **Inventor's Signature:**

Bernard F. Mandrand

3 **Date of Signature:**

30 Jun 1999
Month Day Year

Residence:

Villeurbanne FRANCE FRX
City State or Province Country

Citizenship:

FRANCE

Post Office Address:

21 rue de la Doua

(Insert complete mailing
address, including country)

69100 Villeurbanne FRANCE

1 **Typewritten Full Name
of Joint Inventor**

Given Name Middle Initial Family Name

2 **Inventor's Signature:**

3 **Date of Signature:**

Month Day Year

Residence:

City State or Province Country

Citizenship:

Post Office Address:

(Insert complete mailing
address, including country)

Note to Inventor: Please sign name on line 2 exactly as it appears in line 1 and insert the actual date of signing on line 3.

This form may be executed only when attached to the first page of the Declaration and Power of Attorney of the application to which it pertains.

SEQUENCE LISTING

- (2) INFORMATION FOR SEQ ID NO: 68:
- (i) SEQUENCE CHARACTERISTICS:
- 5 (A) LENGTH: 34 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:
- GACTCGCTGC AGATCGATTT TTTTTTTTTT TTTT 34
- (2) INFORMATION FOR SEQ ID NO: 69:
- (i) SEQUENCE CHARACTERISTICS:
- 15 (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:
- GCCATCAAGC CACCCAAGAA CTCTTAACTT 30
- (2) INFORMATION FOR SEQ ID NO: 70:
- (i) SEQUENCE CHARACTERISTICS:
- 25 (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:
- CCAATAGCCA GACCATTATA TACTACTAATT 30
- (2) INFORMATION FOR SEQ ID NO: 112:
- (i) SEQUENCE CHARACTERISTICS:
- 35 (A) LENGTH: 310 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

```
GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCCTCTGG GAAACCAAGC CCCAGTACTC   60
AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT TTCCTCCCCT CCAGATGGCT   120
AGCCACTGAG GAAGGAAAAA TACTTTCACC TGCAGCTAAC CAACAGAAAT TACTTAAAC   180
CCTTCACCAA ACCTTCCACT TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT   240
TACTGGACCA GGCCTTTTCA AAACATATCA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA   300
AAGAAATAAT                                     310
```

(2) INFORMATION FOR SEQ ID NO: 113:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 103 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```
Leu Ile Glu Gly Pro Leu Val Trp Gly Asn Pro Leu Trp Glu Thr Lys
1           5           10           15
Pro Gln Tyr Ser Ala Gly Lys Ile Glu Xaa Glu Thr Ser Gln Gly His
20           25           30
Thr Phe Leu Pro Ser Arg Trp Leu Ala Thr Glu Glu Gly Lys Ile Leu
35           40           45
Ser Pro Ala Ala Asn Gln Gln Lys Leu Leu Lys Thr Leu His Gln Thr
50           55           60
Phe His Leu Gly Ile Asp Ser Thr His Gln Met Ala Lys Leu Leu Phe
65           70           75           80
Thr Gly Pro Gly Leu Phe Lys Thr Ile Lys Lys Ile Val Arg Gly Cys
85           90           95
Glu Val Cys Gln Arg Asn Asn
100
```

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 635 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```

CCCTGTATCT TTAACCTCCT TGTTAAGTTT GTCTCTTCCA GAATCAAAAC TGTAAACTA 60
CAAATTGTTT TTCAAATGGA GCACCAGATG GAGTCCATGA CTAAGATCCA CCGTGGACCC 120
CTGGACCGGC CTGCTAGCCC ATGCTCCGAT GTTAATGACA TTGAAGGCAC CCCTCCCGAG 180
GAAATCTCAA CTGCACAACC CCTACTATGC CCCAATTCAG CGGGAAGCAG TTAGAGCCGT 240
CATCAGCCAA CCTCCCCAAC AGCACTTGGG TTTTCCTGTT GAGAGGGGGG ACTGAGAGAC 300
AGGACTAGCT GGATTTTCTA GGCCAACGAA GAATCCCTAA GCCTAGCTGG GAAGGTGACT 360
GCATCCACCT CTAAACATGG GGCTTGCAAC TTAGCTCACA CCCGACCAAT CAGAGAGCTC 420
ACTAAAATGC TAATTAGGCA AAAATAGGAG GTAAAGAAAT AGCCAATCAT CTATTGCCTG 480
AGAGCACAGC GGGAGGGACA AGGATCGGGA TATAAACCCA GGCATTGAG CCGGCAACGG 540
CAACCCCTT TGGGTCCCCT CCCTTTGTAT GGGCGCTCTG TTTTCACTCT ATTCACTCT 600
ATTAAATCTT GCAACTGAAA AAAAAAAAAA AAAAA 635

```

(2) INFORMATION FOR SEQ ID NO: 115

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 77 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

```

Pro Cys Ile Phe Asn Leu Leu Val Lys Phe Val Ser Ser Arg Ile Lys
1           5           10           15
Thr Val Lys Leu Gln Ile Val Leu Gln Met Glu His Gln Met Glu Ser
20           25           30
Met Thr Lys Ile His Arg Gly Pro Leu Asp Arg Pro Ala Ser Pro Cys
35           40           45
Ser Asp Val Asn Asp Ile Glu Gly Thr Pro Pro Glu Glu Ile Ser Thr
50           55           60
Ala Gln Pro Leu Leu Cys Pro Asn Ser Ala Gly Ser Ser
65           70           75

```

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleotide

5 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

TGGGGTTCCA TTTGTAAGAC CATCTGTAGC TT 32

10

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1481 base pairs

(B) TYPE: nucleotide

15 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

ATGGCCCTCC CTTATCATAC TTTTCTCTTT ACTGTTCTCT TACCCCTTT CGCTCTCACT 60
GCACCCCTC CATGCTGCTG TACAACCACT AGCTCCCTT ACCAAGAGTT TCTATGAAGA 120
ACGCGGCTTC CTGGAAATAT TGATGCCCA TCATATAGGA GTTATCTAA GGGAACTCC 180
ACCTTCACTG CCCACACCCA TATGCCCGC AACTGCTATA ACTCTGCCAC TCTTTGCATG 240
CATGCAATA CTCATTATTG GACAGGGAAA ATGATTATC CTAGTTGTCC TGGAGGACTT 300
GGAGCCACTG TCTGTTGGAC TTACTTCACC CATACCAGTA TGTCTGATGG GGGTGGAAAT 360
CAAGGTCAGG CAAGAGAAAA ACAAGTAAAG GAAGCAATCT CCCAACTGAC CCGGGGACAT 420
AGCACCCCTA GCCCCTACAA AGGACTAGTT CTCTCAAAAC TACATGAAAC CCTCCGTACC 480
CATACTCGCC TGGTGAGCCT ATTTAATACC ACCCTCACTC GGCTCCATGA GGTCTCAGCC 540
CAAAACCTA CTAAGTGTG GATGTGCCTC CCCCTGCACT TCAGGCCATA CATTTCATC 600
CCTGTTCTTG AACAAATGGA CAACCTCAGC ACAGAAATA ACACCACTTC CGTTTTAGTA 660
GGACCTCTTG TTTCCAATCT GGAATAACC CATACCTCAA ACCTCACCTG TGTAATAATT 720
AGCAATACTA TAGACACAAC CAGCTCCCA TGCATCAGGT GGGTAACACC TCCACACGA 780
ATAGTCTGCC TACCCTCAGG AATATTTTT GTCTGTGGTA CCTCAGCCTA TCATTGTTT 840
AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCTAT GACCATCTAC 900
ACTGAACAAG ATTTATACAA TCATGTCGTA CCTAAGCCCC ACAACAAAAG AGTACCCATT 960
CTTCCTTTTG TTATCAGAGC AGGAGTGCTA GGCAGACTAG GTACTGGCAT TGGCAGTATC 1020

20

ACAACCTCTA CTCAGTTCTA CTACAACTA TCTCAAGAAA TAAATGGTGA CATGGAACAG 1080
 GTCAGTACT CCCTGGTCAC CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT 1140
 CAAAATCGAA GAGCTTTAGA CTTGCTAACC GCCAAAAGAG GGGGAACCTG TTTATTTTAA 1200
 GGAGAAGAAC GCTGTTATTA TGTTAATCAA TCCAGAATTG TCACTGAGAA AGTTAAAGAA 1260
 ATTCGAGATC GAATACAATG TAGAGCAGAG GAGCTTCAAA ACACCGAACG CTGGGGCCTC 1320
 CTCAGCCAAT GGATGCCCTG GGTTCCTCCC TTCTTAGGAC CTCTAGCAGC TCTAATATTG 1380
 TTAATCCTCT TTGGACCCTG TATCTTTAAC CTCCTTGTTA AGTTTGTCTC TTCCAGAATT 1440
 GAAGCTGTAA AGCTACAGAT GGTCTTACAA ATGGAACCCC A 1481

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 493 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

Met Ala Leu Pro Tyr His Thr Phe Leu Phe Thr Val Leu Leu Pro Pro
 1 5 10 15
 Phe Ala Leu Thr Ala Pro Pro Pro Cys Cys Cys Thr Thr Ser Ser Ser
 20 25 30
 Pro Tyr Gln Glu Phe Leu Xaa Arg Thr Arg Leu Pro Gly Asn Ile Asp
 35 40 45
 Ala Pro Ser Tyr Arg Ser Leu Ser Lys Gly Asn Ser Thr Phe Thr Ala
 50 55 60
 His Thr His Met Pro Arg Asn Cys Tyr Asn Ser Ala Thr Leu Cys Met
 65 70 75 80
 His Ala Asn Thr His Tyr Trp Thr Gly Lys Met Ile Asn Pro Ser Cys
 85 90 95
 Pro Gly Gly Leu Gly Ala Thr Val Cys Trp Thr Tyr Phe Thr His Thr
 100 105 110
 Ser Met Ser Asp Gly Gly Gly Ile Gln Gly Gln Ala Arg Glu Lys Gln
 115 120 125
 Val Lys Glu Ala Ile Ser Gln Leu Thr Arg Gly His Ser Thr Pro Ser
 130 135 140
 Pro Tyr Lys Gly Leu Val Leu Ser Lys Leu His Glu Thr Leu Arg Thr

145		150		155		160
His Thr Arg Leu Val Ser Leu Phe Asn Thr Thr Leu Thr Arg Leu His						
	165		170		175	
Glu Val Ser Ala Gln Asn Pro Thr Asn Cys Trp Met Cys Leu Pro Leu						
	180		185		190	
His Phe Arg Pro Tyr Ile Ser Ile Pro Val Pro Glu Gln Trp Asn Asn						
	195		200		205	
Phe Ser Thr Glu Ile Asn Thr Thr Ser Val Leu Val Gly Pro Leu Val						
	210		215		220	
Ser Asn Leu Glu Ile Thr His Thr Ser Asn Leu Thr Cys Val Lys Phe						
225		230		235		240
Ser Asn Thr Ile Asp Thr Thr Ser Ser Gln Cys Ile Arg Trp Val Thr						
	245		250		255	
Pro Pro Thr Arg Ile Val Cys Leu Pro Ser Gly Ile Phe Phe Val Cys						
	260		265		270	
Gly Thr Ser Ala Tyr His Cys Leu Asn Gly Ser Ser Glu Ser Met Cys						
	275		280		285	
Phe Leu Ser Phe Leu Val Pro Pro Met Thr Ile Tyr Thr Glu Gln Asp						
	290		295		300	
Leu Tyr Asn His Val Val Pro Lys Pro His Asn Lys Arg Val Pro Ile						
305		310		315		320
Leu Pro Phe Val Ile Arg Ala Gly Val Leu Gly Arg Leu Gly Thr Gly						
	325		330		335	
Ile Gly Ser Ile Thr Thr Ser Thr Gln Phe Tyr Tyr Lys Leu Ser Gln						
	340		345		350	
Glu Ile Asn Gly Asp Met Glu Gln Val Thr Asp Ser Leu Val Thr Leu						
	355		360		365	
Gln Asp Gln Leu Asn Ser Leu Ala Ala Val Val Leu Gln Asn Arg Arg						
	370		375		380	
Ala Leu Asp Leu Leu Thr Ala Lys Arg Gly Gly Thr Cys Leu Phe Leu						
385		390		395		400
Gly Glu Glu Arg Cys Tyr Tyr Val Asn Gln Ser Arg Ile Val Thr Glu						
	405		410		415	
Lys Val Lys Glu Ile Arg Asp Arg Ile Gln Cys Arg Ala Glu Glu Leu						
	420		425		430	
Gln Asn Thr Glu Arg Trp Gly Leu Leu Ser Gln Trp Met Pro Trp Val						

REPLACEMENT SHEET (RULE 26)

435	440	445
Leu Pro Phe Leu Gly Pro Leu Ala Ala Leu Ile Leu Leu Leu Phe		
450	455	460
Gly Pro Cys Ile Phe Asn Leu Leu Val Lys Phe Val Ser Ser Arg Ile		
465	470	475
Glu Ala Val Lys Leu Gln Met Val Leu Gln Met Glu Pro		480
485	490	

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 32 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CG 32

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 1329 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT GTTTATTTTT	60
AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC ATTACTGAGA AAGTTAAAGA	120
AATTTGAGAT CGAATATAAT GTAGAGCAGA GGACCTTCAA AACACTGCAC CCTGGGGCCT	180
CCTCAGCCAA TGGATGCCCT GGACTCTCCC CTTCTTAGGA CCTCTAGCAG CTATAATATT	240
TTTACTCCTC TTTGGACCT GTATCTTCAA CTTCTTGT AAGTTTGTCT CTTCCAGAAT	300
TGAAGCTGTA AAGCTACAAA TAGTTCTTCA AATGGAACCC CAGATGCAGT CCATGACTAA	360
AATCTACCGT GGACCCCTGG ACCGGCCTGC TAGACTATGC TCTGATGTTA ATGACATTGA	420
AGTCACCCCT CCCGAGGAAA TCTCAACTGC ACAACCCCTA CTACACTCCA ATTCAGTAGG	480
AAGCAGTTAG AGCAGTTGTC AGCCAACCTC CCCAACAGTA CTTGGGTTTT CCTGTTGAGA	540
GGGTGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCNAAGCCT	600

ANCTGGGAAG GTGACCGCAT CCATCTTTAA ACATGGGGCT TGCAACTTAG CTCACACCCG 660
 ACCAATCAGA GAGCTCACTA AAATGCTAAT CAGGCAAAAA CAGGAGGTAA AGCAATAGCC 720
 AATCATCTAT TGCCTGAGAG CACAGCGGGA AGGACAAGGA TTGGGATATA AACTCAGGCA 780
 TTCAAGCCAG CAACAGCAAC CCCCTTTGGG TCCCCTCCCA TTGTATGGGA GCTCTGTTTT 840
 CACTCTATTT CACTCTATTA AATCATGCAA CTGCACTCTT CTGGTCCGTG TTTTTTATGG 900
 CTCAAGCTGA GCTTTTGTTC GCCATCCACC ACTGCTGTTT GCCACCGTCA CAGACCCGCT 960
 GCTGACTTCC ATCCCTTTGG ATCCAGCAGA GTGTCCACTG TGCTCCTGAT CCAGCGAGGT 1020
 ACCCATTGCC ACTCCCGATC AGGCTAAAGG CTTGCCATTG TTCCTGCATG GCTAAGTGCC 1080
 TGGGTTTGTG CTAATAGAAC TGAACACTGG TCACTGGGTT CCATGGTTCT CTTCCATGAC 1140
 CCACGGCTTC TAATAGAGCT ATAACACTCA CCGCATGCCC CAAGATTCCA TTCCTTGTA 1200
 TCTGTGAGGC CAAGAACCCC AGGTCAGAGA ANGTGAGGCT TCCCACCATT TGGGAAGTGG 1260
 CCCACTGCCA TTTTGGTAGC GGGCCACCAC CATCTTGGGA GCTGTGGGAG CAAGGATCCC 1320
 CCAGTAACA 1329

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

Gln	Asn	Arg	Arg	Ala	Leu	Asp	Leu	Leu	Thr	Ala	Lys	Arg	Gly	Gly	Thr
1				5					10					15	
Cys	Leu	Phe	Leu	Gly	Glu	Glu	Cys	Cys	Xaa	Tyr	Val	Asn	Gln	Ser	Gly
			20						25					30	
Ile	Ile	Thr	Glu	Lys	Val	Lys	Glu	Ile	Xaa	Asp	Arg	Ile	Xaa	Cys	Arg
			35						40					45	
Ala	Glu	Asp	Leu	Gln	Asn	Thr	Ala	Pro	Trp	Gly	Leu	Leu	Ser	Gln	Trp
			50						55					60	
Met	Pro	Trp	Thr	Leu	Pro	Phe	Leu	Gly	Pro	Leu	Ala	Ala	Ile	Ile	Phe
65						70					75				80
Leu	Leu	Leu	Phe	Gly	Pro	Cys	Ile	Phe	Asn	Phe	Leu	Val	Lys	Phe	Val
			85								90				95
Ser	Ser	Arg	Ile	Glu	Ala	Val	Lys	Leu	Gln	Ile	Val	Leu	Gln	Met	Glu
							100								110

Pro Gln Met Gln Ser Met Thr Lys Ile Tyr Arg Gly Pro Leu Asp Arg
 115 120 125
 Pro Ala Arg Leu Cys Ser Asp Val Asn Asp Ile Glu Val Thr Pro Pro
 130 135 140
 Glu Glu Ile Ser Thr Ala Gln Pro Leu Leu His Ser Asn Ser Val Gly
 145 150 155 160
 Ser Ser

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

10 GGCATTGATA GCACCCATCA G 21

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

20 CATGTCACCA GGGTGAATA G 21

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 758 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

GGCATTCATA GCACCCATCA GATGGCCAAA TCATTATTTA CTGGACCAGG CCTTTTCAAA 60
ACTATCAAGC AGATAGGGCC CGTGAAGCAT GCCAAAGAAA TAATCCCCTG CCTTATCGCC 120
ATGTTCCCTC AGGAGAACAA AGAACAGGCC ATTACCCAGG GGAAGACTGG CAACTAGATT 180
TTACCCACAT GGCCAAATGT CAGGGATTTC AGCATCTACT AGTCTGGGCA GATACTTTCA 240
CTGGTTGGGT GGAGTCTTCT CCTTG TAGGA CAGAAAAGAC CCAAGAGGTA ATAAAGGCAC 300
TAATGAAATA ATTCCACAT TTGGACTTCC CCCAGGATTA CAGGGTGACA ATGGCCCCGC 360
TTTCAAGGCT GCAGTAACCC AGGGACTATC CCAGGTGTTA GCCATACAAT ATCACTTACA 420
CTGTGCCTGG AGGCCACAAT CCTCCAGAAA AGTCAAGAAA ATGAATGAAA CACTCAAAGA 480
TCTAAAAAAG CTAACCCAAG AAACCCACAT TGCATGACCT GTTCTGTTGC CTATAACCTT 540
ACTAAGAATC CATAACTATC CCCCAAAAG CAGGACTTAG CCCATACGAG ATGCTATATG 600
GATGGCCTTT CCTAACCAAT GACCTTGTC TTGACTGAGA AATGGCCAAC TTAGTTGCAG 660
ACATCACCTC CTTAGCCAAA TATCAACAAG TTCTTAAAC ATCACAGGGA ACCTGTCCCC 720
GAGAGGAGGG AAAGGAACTA TTCCACCCTG GTGACATG 758

10 (2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

CGGACATCCA AAGTGATGGG AAACG

25

20 (2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

GGACAGGAAA GTAAGACTGA GAAGGC

26

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 26 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

CCTAGAACGT ATTCTGGAGA ATTGGG

26

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 26 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

TGGCTCTCAA TGGTCAAACA TACCCG

26

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 1511 base pairs
(B) TYPE: nucleotide
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG 60

```

ATTATATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT CTTCAAGGGA GAGAAACCTG 120
GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA 180
GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACTTTC TTTTCATTAA GAGACAACTC 240
ACAATTATGT AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA 300
CCCCAGCGTC CCCTCCCCGA CTCCTTCCTC AACTAATAAG GACCCCCCTT TAACCCAAAC 360
GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA AAGAGTGCCA ATATTCCCCG 420
ATTATGCCCC CTCCAAGCAG TCAGAGGAGG AGAATTCGGC CCAGCCAGAG TGCCTGTACC 480
TTTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAAATTCT CAGATAACCC 540
TGACCGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA 600
TATAATGTTA CTACTAAATC AGACACTAAC CCCAAATGAG AGAAGTGCCG CTGTAACCTG 660
AGCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAATAGGA TGACAACAGA 720
GGAAGAACA ACTCCACAG GCCAGCAGGC AGTTCACAGT GTAGACCCTC ATTGGGACAC 780
AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTGCTA ACTTGCGTGC TAGAAGGACT 840
GAGGAAAAC AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA 900
GGAAGAAAAT CTTACTGCTT TTCTCGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC 960
CCTGTCACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTTA TCACTCAGTC 1020
AGCTGCAGAC ATTAGAAAAA ACTTCAAAAG TCTGCCTTAG GCGCGGAGCA GAACTTAGAA 1080
ACCCTATTTA ACTTGGCATC CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCGAAA 1140
CGGGACAAAC GGGATAAAAA AAAAAGGGGG GGTCCACTAC TTTAGTCATG GCGCTCAGGC 1200
AAGCAGACTT TGGAGGCTCT GCAAAGGGA AAAGCTGGGC AAATCAAATG CCTAATAGGG 1260
CTGGCTTCCA GTGCGGTCTA CAAGGACACT TAAAAAGA TTATCCAAGT AGAAATAAGC 1320
CGCCCCCTTG TCCATGCCCC TTACGTCAAG GGAATCACTG GAAGGCCAC TGCCCCAGGG 1380
GATGAAGATA CTCTGAGTCA GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT 1440
GCCCCGGGCG AGCGCCAGCC CATGCCATCA CCCTCACAGA GCCCCGGGTA TGTTTGACCA 1500
TTGAGAGCCA A 1511

```

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 352 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

Leu Glu Arg Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu

10

1

5

10

15

Arg	Lys	Lys	Arg	Phe	Ile	Phe	Phe	Cys	Ser	Thr	Ala	Trp	Pro	Gln	Tyr			
			20					25					30					
Pro	Leu	Gln	Gly	Arg	Glu	Thr	Trp	Leu	Pro	Glu	Gly	Ser	Ile	Asn	Tyr			
		35					40					45						
Asn	Ile	Ile	Leu	Gln	Leu	Asp	Leu	Phe	Cys	Arg	Lys	Glu	Gly	Lys	Trp			
		50					55					60						
Ser	Glu	Val	Pro	Tyr	Val	Gln	Thr	Phe	Phe	Ser	Leu	Arg	Asp	Asn	Ser			
		65				70				75				80				
Gln	Leu	Cys	Lys	Lys	Cys	Gly	Leu	Cys	Pro	Thr	Gly	Ser	Pro	Gln	Ser			
				85					90					95				
Pro	Pro	Pro	Tyr	Pro	Ser	Val	Pro	Ser	Pro	Thr	Pro	Ser	Ser	Thr	Asn			
			100					105						110				
Lys	Asp	Pro	Pro	Leu	Thr	Gln	Thr	Val	Gln	Lys	Glu	Ile	Asp	Lys	Gly			
			115				120						125					
Val	Asn	Asn	Glu	Pro	Lys	Ser	Ala	Asn	Ile	Pro	Arg	Leu	Cys	Pro	Leu			
			130				135						140					
Gln	Ala	Val	Arg	Gly	Gly	Glu	Phe	Gly	Pro	Ala	Arg	Val	Pro	Val	Pro			
		145				150				155				160				
Phe	Ser	Leu	Ser	Asp	Leu	Lys	Gln	Ile	Lys	Ile	Asp	Leu	Gly	Lys	Phe			
			165					170					175					
Ser	Asp	Asn	Pro	Asp	Gly	Tyr	Ile	Asp	Val	Leu	Gln	Gly	Leu	Gly	Gln			
			180					185					190					
Ser	Phe	Asp	Leu	Thr	Trp	Arg	Asp	Ile	Met	Leu	Leu	Leu	Asn	Gln	Thr			
			195				200						205					
Leu	Thr	Pro	Asn	Glu	Arg	Ser	Ala	Ala	Val	Thr	Ala	Ala	Arg	Glu	Phe			
			210				215						220					
Gly	Asp	Leu	Trp	Tyr	Leu	Ser	Gln	Ala	Asn	Asn	Arg	Met	Thr	Thr	Glu			
		225				230				235				240				
Glu	Arg	Thr	Thr	Pro	Thr	Gly	Gln	Gln	Ala	Val	Pro	Ser	Val	Asp	Pro			
			245					250					255					
His	Trp	Asp	Thr	Glu	Ser	Glu	His	Gly	Asp	Trp	Cys	His	Lys	His	Leu			
			260					265					270					
Leu	Thr	Cys	Val	Leu	Glu	Gly	Leu	Arg	Lys	Thr	Arg	Lys	Lys	Pro	Met			
			275				280					285						
Asn	Tyr	Ser	Met	Met	Ser	Thr	Ile	Thr	Gln	Gly	Lys	Glu	Glu	Asn	Leu			
			290				295					300						

1	5					10					15					
Arg	Gly	Ser	His	Met	Ala	Ser	Met	Thr	Gly	Gly	Gln	Gln	Met	Gly	Arg	
20					25					30						
Ile	Leu	Glu	Arg	Ile	Leu	Glu	Asn	Trp	Asp	Gln	Cys	Asp	Thr	Gln	Thr	
35					40					45						
Leu	Arg	Lys	Lys	Arg	Phe	Ile	Phe	Phe	Cys	Ser	Thr	Ala	Trp	Pro	Gln	
50					55					60						
Tyr	Pro	Leu	Gln	Gly	Arg	Glu	Thr	Trp	Leu	Pro	Glu	Gly	Ser	Ile	Asn	
65					70					75					80	
Tyr	Asn	Ile	Ile	Leu	Gln	Leu	Asp	Leu	Phe	Cys	Arg	Lys	Glu	Gly	Lys	
85					90					95						
Trp	Ser	Glu	Val	Pro	Tyr	Val	Gln	Thr	Phe	Phe	Ser	Leu	Arg	Asp	Asn	
100					105					110						
Ser	Gln	Leu	Cys	Lys	Lys	Cys	Gly	Leu	Cys	Pro	Thr	Gly	Ser	Pro	Gln	
115					120					125						
Ser	Pro	Pro	Pro	Tyr	Pro	Ser	Val	Pro	Ser	Pro	Thr	Pro	Ser	Ser	Thr	
130					135					140						
Asn	Lys	Asp	Pro	Pro	Leu	Thr	Gln	Thr	Val	Gln	Lys	Glu	Ile	Asp	Lys	
145					150					155					160	
Gly	Val	Asn	Asn	Glu	Pro	Lys	Ser	Ala	Asn	Ile	Pro	Arg	Leu	Cys	Pro	
165					170					175						
Leu	Gln	Ala	Val	Arg	Gly	Gly	Glu	Phe	Gly	Pro	Ala	Arg	Val	Pro	Val	
180					185					190						
Pro	Phe	Ser	Leu	Ser	Asp	Leu	Lys	Gln	Ile	Lys	Ile	Asp	Leu	Gly	Lys	
195					200					205						
Phe	Ser	Asp	Asn	Pro	Asp	Gly	Tyr	Ile	Asp	Val	Leu	Gln	Gly	Leu	Gly	
210					215					220						
Gln	Ser	Phe	Asp	Leu	Thr	Trp	Arg	Asp	Ile	Met	Leu	Leu	Leu	Asn	Gln	
225					230					235					240	
Thr	Leu	Thr	Pro	Asn	Glu	Arg	Ser	Ala	Ala	Val	Thr	Ala	Ala	Arg	Glu	
245					250					255						
Phe	Gly	Asp	Leu	Trp	Tyr	Leu	Ser	Gln	Ala	Asn	Asn	Arg	Met	Thr	Thr	
260					265					270						
Glu	Glu	Arg	Thr	Thr	Pro	Thr	Gly	Gln	Gln	Ala	Val	Pro	Ser	Val	Asp	
275					280					285						
Pro	His	Trp	Asp	Thr	Glu	Ser	Glu	His	Gly	Asp	Trp	Cys	His	Lys	His	

290	295	300
Leu Leu Thr Cys Val Leu Glu Gly Leu Arg Lys Thr Arg Lys Lys Pro		
305	310	315 320
Met Asn Tyr Ser Met Met Ser Thr Ile Thr Gln Gly Lys Glu Glu Asn		
325	330	335
Leu Thr Ala Phe Leu Asp Arg Leu Arg Glu Ala Leu Arg Lys His Thr		
340	345	350
Ser Leu Ser Pro Asp Ser Ile Glu Gly Gln Leu Ile Leu Lys Asp Lys		
355	360	365
Phe Ile Thr Gln Ser Ala Ala Asp Ile Arg Lys Asn Phe Lys Ser Leu		
370	375	380
Pro Lys Leu Ala Ala Ala Leu Glu His His His His His His		
385	390	395

(2) INFORMATION FOR SEQ ID NO: 137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 378 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Leu Glu Arg		
1	5	10 15
Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu Arg Lys Lys		
20	25	30
Arg Phe Ile Phe Phe Cys Ser Thr Ala Trp Pro Gln Tyr Pro Leu Gln		
35	40	45
Gly Arg Glu Thr Trp Leu Pro Glu Gly Ser Ile Asn Tyr Asn Ile Ile		
50	55	60
Leu Gln Leu Asp Leu Phe Cys Arg Lys Glu Gly Lys Trp Ser Glu Val		
65	70	75 80
Pro Tyr Val Gln Thr Phe Phe Ser Leu Arg Asp Asn Ser Gln Leu Cys		
85	90	95
Lys Lys Cys Gly Leu Cys Pro Thr Gly Ser Pro Gln Ser Pro Pro Pro		
100	105	110

Tyr	Pro	Ser	Val	Pro	Ser	Pro	Thr	Pro	Ser	Ser	Thr	Asn	Lys	Asp	Pro
			115				120					125			
Pro	Leu	Thr	Gln	Thr	Val	Gln	Lys	Glu	Ile	Asp	Lys	Gly	Val	Asn	Asn
			130				135					140			
Glu	Pro	Lys	Ser	Ala	Asn	Ile	Pro	Arg	Leu	Cys	Pro	Leu	Gln	Ala	Val
			145				150				155				160
Arg	Gly	Gly	Glu	Phe	Gly	Pro	Ala	Arg	Val	Pro	Val	Pro	Phe	Ser	Leu
				165					170					175	
Ser	Asp	Leu	Lys	Gln	Ile	Lys	Ile	Asp	Leu	Gly	Lys	Phe	Ser	Asp	Asn
			180						185					190	
Pro	Asp	Gly	Tyr	Ile	Asp	Val	Leu	Gln	Gly	Leu	Gly	Gln	Ser	Phe	Asp
			195					200				205			
Leu	Thr	Trp	Arg	Asp	Ile	Met	Leu	Leu	Leu	Asn	Gln	Thr	Leu	Thr	Pro
			210					215				220			
Asn	Glu	Arg	Ser	Ala	Ala	Val	Thr	Ala	Ala	Arg	Glu	Phe	Gly	Asp	Leu
			225				230				235				240
Trp	Tyr	Leu	Ser	Gln	Ala	Asn	Asn	Arg	Met	Thr	Thr	Glu	Glu	Arg	Thr
				245					250					255	
Thr	Pro	Thr	Gly	Gln	Gln	Ala	Val	Pro	Ser	Val	Asp	Pro	His	Trp	Asp
			260						265					270	
Thr	Glu	Ser	Glu	His	Gly	Asp	Trp	Cys	His	Lys	His	Leu	Leu	Thr	Cys
			275					280					285		
Val	Leu	Glu	Gly	Leu	Arg	Lys	Thr	Arg	Lys	Lys	Pro	Met	Asn	Tyr	Ser
			290					295				300			
Met	Met	Ser	Thr	Ile	Thr	Gln	Gly	Lys	Glu	Glu	Asn	Leu	Thr	Ala	Phe
			305				310				315				320
Leu	Asp	Arg	Leu	Arg	Glu	Ala	Leu	Arg	Lys	His	Thr	Ser	Leu	Ser	Pro
				325					330				335		
Asp	Ser	Ile	Glu	Gly	Gln	Leu	Ile	Leu	Lys	Asp	Lys	Phe	Ile	Thr	Gln
			340						345				350		
Ser	Ala	Ala	Asp	Ile	Arg	Lys	Asn	Phe	Lys	Ser	Leu	Pro	Lys	Leu	Ala
			355					360					365		
Ala	Ala	Leu	Glu	His	His	His	His	His	His	His	His	His	His	His	His
			370					375							

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

CTTGGAGGGT GCATAACCAG GGAAT

25

5

(2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleotide

10 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

TGTCCGCTGT GCTCCTGATC

20

15

(2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleotide

20 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

CTATGTCCTT TTGGACTGTT TGGGT

25

25

(2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 764 base pairs

(B) TYPE: nucleotide

30 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

```

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT GGGCTAAAGG 60
CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTTCATC CTAATCGAGC TGAACACTAG 120
TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180
CTGCATGGTC CAAGATTCCA TTCCTTGGAA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA 240
ACACAAGGCT TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGAAGC AGCCCGCCAC 300
TATCTTGGGA GCTCTGGGAG CAAGGACCCC AGGTAACAAT TTGGTGACCA CGAAGGGACC 360
TGAATCCGCA ACCATGAAGG GATCTCCAAA GCAATTGGAA ATGTTCTCTC CAAGGCCAAA 420
ATGCCCCATA GATGTATTCT GGAGAATTGG GACCAATTG ACCCTCAGAC AGTAAGAAAA 480
AAATGACTTA TATTCTTCTG CAGTACCGCC CTGGCCACGA TATCCTCTTC AAGGGGGAGA 540
AACCTGGCCT CCTGAGGGAA GTATAAATTA TAACACCATC TTACAGCTAG ACCTGTTTTG 600
TAGAAAAGGA GGCAAATGGA GTGAAGTGCC ATATTACAA ACTTTCTTTT CATTAAAAGA 660
CAACTCGCAA TTATGTAAAC AGTGTGATT GTGTTCCTAC ACGGAAGCCC TCAGATTCTA 720
CTCCCCACCC CCGGCATCTC CCCTGAATCC CTCCCCAACT TATT 764

```

(2) INFORMATION FOR SEQ ID NO: 142:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 800 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

```

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT GGGCTAAAGG 60
CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTTCATC CTAATCGAGC TGAACACTAG 120
TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180
CTGCATGGTC CAAGATTCCA TTCCTTGGAA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA 240
ACACAAGGCT TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGAAGC GGCCCGCCAC 300
TATCTTGGGA GCTCTGGGAG CAAGGACCCC CAGGTAACAA TTTGGTGACC ACGAAGGGAC 360
CTGAATCCGC AACCATGAAG GGATCTCCAA AGCAATTGGA AATGTTCTCT CCAAGGCCAAA 420
AATGCCCTTA AGATGTATTG TGGAGAATTG GGACCAATCT GACCCTCAGA CAGTAAGAAA 480
AAAAATGACT TATATTCTTC TGCAGTACCG CCTGGCCACG GATATCCTCT TCAAGGGGGA 540
GAAACCTGGC CTCCTGAGGG AAGTATAAAT TATAACACCA TCTTACAGCT AGACCTGTTT 600
TGTAAGAAAG GAGGCAAATG GAGTGAAGTG CCATATTTAC AAACCTTCTT TTCATTAAAA 660
GACAACTCGC AATTATGTAA ACAGTGTGAT TTGTGTCCTA CAGGAAGCCC TCAGATCTAC 720
CTCCCTACCC CCGCATCTCC CTGACTCCTT CCCCAACTAA TAAGGACCCA CTTGAGCCCA 780
AACAGTCCAA AAGGACATAG 800

```